DYNAMIC-BASED STRUCTURE MEASURES OF COMPLEX NETWORKS

ZHU GUIMEI

NATIONAL UNIVERSITY OF SINGAPORE

2012

DYNAMIC-BASED STRUCTURE MEASURES OF COMPLEX NETWORKS

ZHU GUIMEI

(M.Sc., University of Science and Technology of China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

NUS GRADUATE SCHOOL FOR INTEGRATIVE

SCIENCES AND ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2012

Dedicated to My Parents and My Loved Ones

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.
This thesis has also not been submitted for any degree in any university previously.

Zhu Guimei Zhu Guimei

10 Aug 2012 10 Aug 2012

DYNAMIC-BASED STRUCTURE MEASURES of COMPLEX NETWORKS

NUS Graduate School for Integrative Sciences and Engineering

Block S16, Level 8, 6 Science Drive 2

National University of Singapore

Singapore 117546

Email: zhugm07@gmail.com

Acknowledgements

It is one of the most precious and fruitful times in my life to do research and pursue my PhD in National University of Singapore. NUS witnesses my growth not only on research but also my life. In this period, I met, learned from and get along well with many good friends and mentors. I want to thank all of them from the bottom of my heart for their always warm help and gracious support.

First and foremost, I would like to express my sincere thanks to Professor Li Baowen. In the progress of working with him, I learned quite a lot, not only his research approaches but also his attitude toward research. He is quite strict to the research, before finishing any work, he must think it in a systematic way, to check its novelty and significance. And also he is quite open mind to collaborate with and seek for comments from others. Actually in the progress of doing my research, I always feel unconfident when I compared myself to other very outstanding cohorts and friends, However, Prof Li's constant support, encouragement, and instructive guidance helped me cheer up, grown up, and eventually made my Ph. D thesis.

Meanwhile, I am extremely grateful to Professor Chen Yuzong, who dedicate most of his time to his students and research. His concentration and diligence on his research field quite impressed me. I appreciate his always warm guidance and support very much.

My hearted thanks go to my USTC senior, and also collaborator Professor Yang Huijie who guides me to do the research step by step. Also special thanks to Professor Sarika Jalan, in the progress of research, she taught me how to do it independently. I learnt and grown up quite a lot from collaborating with her. Thanks also goes to Professor Lai Ying-Chen, his concentration on work and his high efficient work impress me a lot. Thanks to all of their guidance, patience and numerous discussions.

Many thanks to Professor Peter Hänggi, who is always with unlimited energy and passion toward research. And from his hard working attitude, I understand that harvest not occasionally happen, it always goes to prepared people.

I also want to give my sincere thanks to all the professors help me for my module and research: Professors Song Jianxing, Gong Jiangbin, Wang Jiansheng, Hu Bambi, Liu Zonghua, Wu Changqin, Wu Gang, Zhang Gang, Zhao Ming, Wang Wenxu, Huang Liang and etc.

Friendship is always my spiritual support, I would like to sincerely thank my close best friends, the sweet couple of Fang Chunliu and Chen Jie , Cao Ye and Liu Sha, the lovely young and quite mature Qin Chu, always A+ student Yang Lina, considerate, floral fan Tao Lin, and good Hou Ruizheng, especially they helped and encouraged me go through the quite tough time in 2009 and 2013.

I also thanks my many other good friends: Zhang Lifa and Zhang Congmei, sweet Zhang Kaiwen, Feng Ling, Tang Qinglin, Liu Dan, Qiao Zhi, Xu Wen, Tang Yunfei, Wang lei, Lan Jinhua, Li Nianbei and Zhang Lei ping, Yang Nuo, Dario Poletti, Yao donglai, Ren jie, Ni Xiaoxi, Shi Lihong, Tinh, Zhang Xun, Ma Jing, Zhao Xiangming, Xu Xiangfan, Xie Rongguo, Yang Rui, Wang Chen and etc.

Last but not the least, I would like to express my deepest thanks to my parents

and my younger brother. They always stand there and prepare to help and support me whenever I need them. They are always my strong backup force for me. Family means everything to me. I just want to express my heartiest thanks to them, and I love my family very much!

Table of Contents

Abstract

Complex networks research has attracted incredible wide attention in recent years. Although great progress has been achieved, the measure of complex networks is not yet fully understood. We still do not have any systematic program for characterizing network structures. Furthermore, the measures of network structures, such as the microproperties, the patterns at different scales, and the macroproperties, are generally simple applications of the concepts in graph theory, bioinformatics, social science, and fractal theory. They are not dynamics based.

We cannot expect simple and reasonable relations between the structure measures and the dynamical processes on networks. The lack of powerful tools to characterize network structures is an essential bottleneck in understanding dynamical processes on networks. Hence the general aim of this dissertation is to excavate the keys to these problems by the way using the dynamic-based structure measures of complex networks. The structures of complex networks can induce nontrivial properties in the physical processes occurring on them. The physical processes in turn can be used as a probe to capture the structural properties.

To understand the localization properties of complex networks, the methods

from Random Matrix Theory(RMT) and dynamics systems are demonstrated in the first place. For undirected networks, the structural characteristics of complex networks through the representative eigenvectors of the adjacent matrix are studied. The probability distribution functions of the components of the representative eigenvectors are proposed to describe the localization on networks where the Euclidean distance is invalid. Several quantities are also used to describe the localization properties of the representative states, such as the participation ratio, the structural entropy, and the probability distribution function of the nearest neighbor level spacings for spectra of complex networks. Whole-cell networks in the real world and the Watts-Strogatz small-world and Barabasi-Albert scale-free networks are considered.

Then for the directed networks with inhibitory and excitatory couplings spectra analysis, the particular eigenvector localization properties of random networks for different values of correlation among their entries are investigated. Spectra of random networks with completely uncorrelated entries show a circular distribution with delocalized eigenvectors, whereas networks with correlated entries have localized eigenvectors. In order to understand the origin of localization, the spectra as a function of connection probability and directionality were traced here. As connections are made directed, eigenstates start occurring in complex-conjugate pairs and the eigenvalue distribution combined with the localization measure shows a rich pattern.

The studies of spectra analysis of the system level dynamics of these large scale networks provided that the networks have remarkable localization properties due to the nontrivial topological structures, and the ascending-order-ranked series

of the occurrence probabilities at the nodes behave generally multi-fractal. It can be used as a structural measure of complex networks. The study also leads to some significant insights into the evolutionary process underpinning the networks.

The last but not least topic is to detect the evolutionary history of a network. It was noticed that in a complex network, different groups of nodes may have existed for different amounts of time. To detect the evolutionary history of a network is of great importance. A spectral-analysis based method to address this fundamental question in network science is presented here. In particular, it was found that there are complex networks in the real-world for which there is a positive correlation between the eigenvalue magnitude and node age. It should be noted, however, that at the present the applicability of our method is limited to the networks for which information about the node age has been encoded gradually in the eigen-properties through evolution.

List of Publications

[1] **Guimei Zhu**, Huijie Yang, Rui Yang, Jie Ren, Baowen Li, and Ying-Cheng Lai, "Uncovering Evolutionary Ages of Nodes in Complex Networks", European Physical Journal B, **85**, 106 (2012).

[2] Sarika Jalan, **Guimei Zhu**, and Baowen Li, "Spectral Properties of Directed Random Networks with Modular Structure", Physical Review E, **84**, 046107 (2011).

[3] **Guimei Zhu**, and Baowen Li, "Phonons on Complex Networks", PHONONICS 2011 International Conference on PHONONIC Crystals, Metamaterials & Optomechanics, **1**, 154 (2011).

[4] **Guimei Zhu**, Huijie Yang, Chuanyang Yin, and Baowen Li, "Localizations on Complex Networks", Physical Review E, **77**, 066113 (2008).

[5] Huijie Yang, Chuanyang Yin, **Guimei Zhu**, and Baowen Li, "Self-affine fractals embedded in spectra of complex networks", Physical Review E, **77**, 045101 **R** (2008).

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivations

Recent years have witnessed tremendous research activities to the complex network and its applications in diverse fields [1]. Structural measures of complex networks are the cornerstone to understand relations between the structures, dynamics and functions. Real world networks generally have nontrivial properties as the small-world [2], scale-free [3], motif [4], modularity, hierarchy [5], fractal [6] and so on. The small-world effect is that in average the nodes can reach each other with only a small number of hops. The scale-free refers to the number of edges per node obeys a right-skewed distribution. It is also found that some special subgraphs containing several connected nodes, called motifs, occur with significant probabilities compared with that in the corresponding randomized networks. These three individual, pair or local pattern-based properties are called micro-properties. On the other hand, the modularity is a kind of macro-property represents that a network can be separated into loosely connected groups within which the nodes are

densely connected, respectively.

To a certain degree, dynamics on networks can be regarded as the transport processes of mass, energy, signal and/or information at different structure scales [7, 8]. Sometimes we have to deal with networks with unreasonable large number of nodes and edges, e.g., the neuron networks and the World-Wide-Web networks, when designing a coarse-grain procedure is essential [9]. The patterns at different scales may provide a reasonable solution to these problems. It is found that some real world networks have hierarchical structures, in which the small-world and the scale-free properties can coexist [5]. Moreover, many real world networks behave self-similar at different structure levels (fractal) [6] .

Though great progresses have been archived, the measures of complex networks are not yet fully understood. Just as pointed out by Newman [10], that our techniques for analyzing networks are at present time no more than a grab-bag of miscellaneous and largely unrelated tools, and we still do not have a systematic program for characterizing network structures. Furthermore, the measures of network structures, such as the micro-properties, the patterns at different scales and the macro-properties, are generally a simple application of the concepts in graph theory, bioinformatics, social science and fractal theory, namely, they are not dynamic-based. We can not expect simple and reasonable relations between the measures and the dynamical processes on networks.

The lack of powerful tools to characterize network structures is an essential bottleneck to understand dynamical processes on networks. One typical example is the synchronizabilities of complex networks. Detailed works show that almost all the structure measures affect the synchronizabilities in complicated ways [11], based upon which we can not reach a clear picture of the mechanisms for synchronization

processes on networks.

Dynamic-based measures of complex networks may be the key to the problems. The structures of complex networks can induce nontrivial properties to the physical processes occurring on them. The physical processes in turn can be used as the probe to capture the structure properties. Well studied dynamical processes, such as the random walks [12, 13] and the Boolean dynamics [14], can be good candidates as probes. To cite an example, the random walks on complex networks that biased towards a target node show a localization-delocalization transition [12].

In this thesis, we would like to study the large scale structure and systematic level dynamics of certain model networks and real world networks using tools from Random Matrix Theory(RMT)[15] and nonlinear dynamics. Our studies of spectra analysis of the systematic level dynamics of these large scale networks provide that the networks have remarkable localization properties due to the nontrivial topological structures, and the ascending-order-ranked series of the occurrence probabilities at the nodes behave generally multi-fractal. It can be used as a dynamic-based structural measure of complex networks. Our study also leads to some significant insights into the evolutionary process underpinning the networks.

To make the thesis self-consistent, I would like to introduce the general concepts of complex networks in Section 1.2, and also give the explanation of dynamic-based structure measure of complex network in Section 1.3.

1.2 General Description of Complex Networks

Complex networks are all around us, examples consist of the Internet, the World Wide Web, airline and transportation networks, electric power grids, social

networks of acquaintance, collaboration networks, neural networks, protein-protein networks, metabolic networks, food webs, distribution networks such as blood vessels or postal delivery routes, networks of citations between papers, and many others (Fig.1.1 from Ref. [10]). In the mean times, ourselves, as individuals, are the cells of various social relationship networks.

Recent years have witnessed a substantial new movement in network researches[19], with the attention shift away from the study of single small graphs to the large scale statistical graphs. This storm of activities, were stirred up by two seminal papers, one by Watts and Strogatz on small-world networks [2], and another one by Barabsi and Albert on scale-free networks [3](Barabasi 1999), has been driven largely by the possibility of computers and communication networks which allow us to gather and analyze data on an unbelievable large scale than before.

For the complex network research, we notice that the network structure has crucial consequences on the network functional robustness and response to external perturbations, as random failures, or targeted attacks. Meanwhile, it helps to study the dynamical behavior of large interacting. This led to a series of evidences pointing to the crucial role played by the network topology in determining the emergence of collective dynamical behavior, such as synchronization, or in governing the main features of relevant processes that take place in complex networks, such as the spreading of epidemics, information and rumors. So structure is the cornerstone for understanding the relationship between structures, function, dynamics of complex networks [10] [20][21].

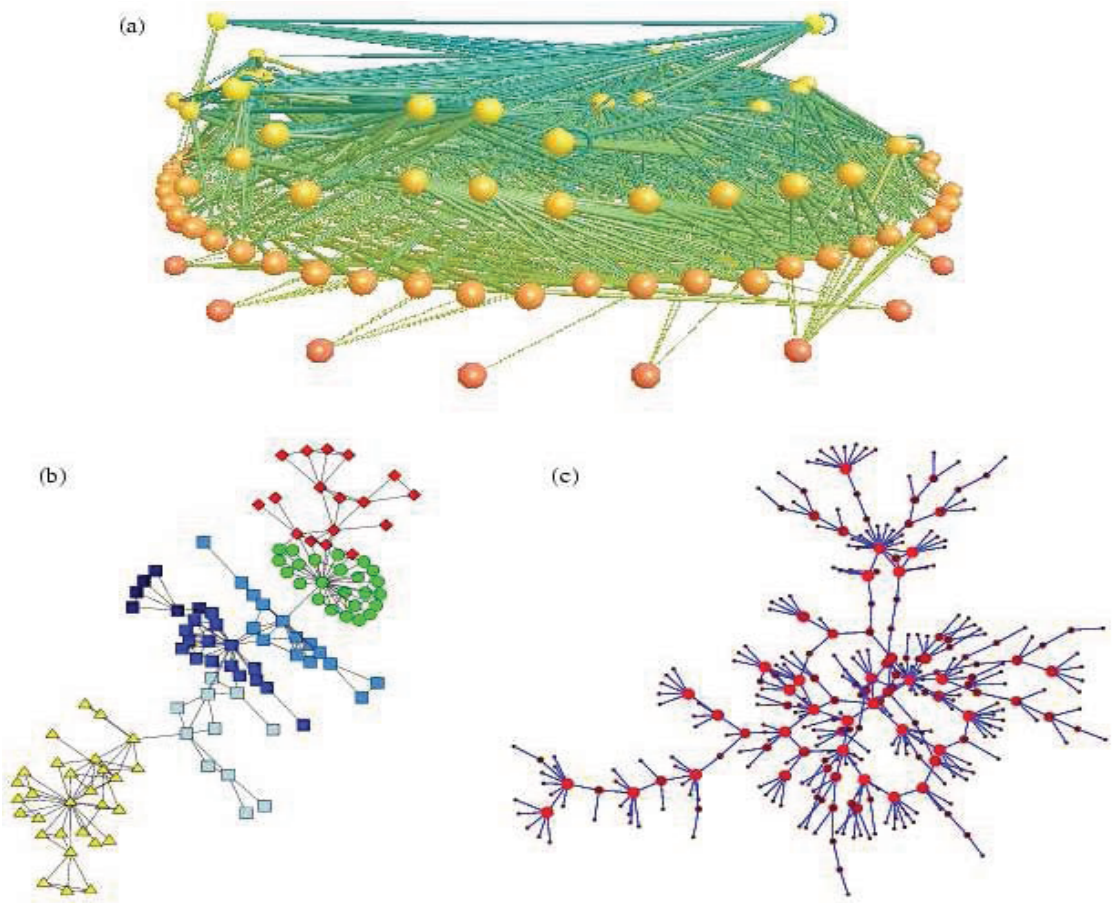

Figure 1.1: **Three examples of complex networks in the real world.** (a) A food web of predator-prey interactions between species in a freshwater lake [16]. (b) The network of collaborations between scientists at a private research institution [17]. (c) A network of sexual contacts between individuals in the study by Potterat et al. [18]. Adapted from Ref. [10].

1.2.1 Basic Concepts in Complex Networks

In the study of networks, the network size is a set of N nodes. The degree of a node in a network (sometimes called connectivity) is the number of connections or edges the node has to other nodes.

In a undirected graph, each of the edges is defined by a couple of nodes i and j , and is denoted as (i, j) or l_{ij}. Usually, we draw a dot for each node and join two dots by an edge to have a graph. Note that the picture does not allow self-connecting nodes or multiple edges. For a directed network, edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges. A weighted network is a network where the edges among nodes have weights assigned to them. (see Fig. 1.2 from Ref.[21])

For a network G of size N, the number of edges D is at least 0 and at most $\frac{N\cdot(N-1)}{2}$ (when all the nodes are pair-wise adjacent). G is supposed to be sparse for real world networks.

Considering a matricial representation of a network, a network $G = (N)$ can be described by giving the adjacency matrix A, a $N \times N$ square matrix whose elements $A_{ij}(i, j = 1, \cdots, N)$ are equal to 1 when the edge l_{ij} exists, and zero otherwise. The diagonal of the adjacency matrix contains zeros. Therefore the matrix is a symmetric one for undirected graphs.

The Degree k_i of the node i is the number of edges connected with the node, and is defined in terms of the adjacency matrix A as,

$$K_i = \sum_{j \in n} A_{ij}. \tag{1.1}$$

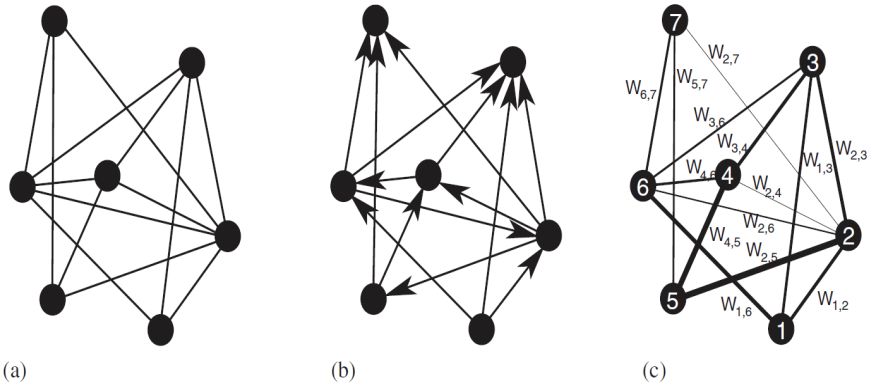

Figure 1.2: **Graphical representation of a undirected (a), a directed (b), and a weighted undirected (c) graph.** These networks are all with $N = 7$ nodes and $K = 14$ links. In the directed graph, adjacent nodes are connected by arrows, indicating the direction of each link. In the weighted graph, the values $\omega_{i,j}$ reported on each link indicate the weights of the links, and are graphically represented by the link thicknesses. Adapted from Ref. [21].

Degree distribution $P(k)$ is the most basic topological characterization of a network G. It is defined as the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have

$$P_k = \frac{n_k}{n}. \tag{1.2}$$

The degree distribution is very important in studying both model networks and real networks, such as the Internet and social networks, and theoretical networks. The simplest network model, for example, the (Bernoulli) random graph, in which each of n nodes is connected (or not) with independent probability $p(or 1-p)$, has a binomial distribution of degrees (or Poisson in the limit of large n). Most networks in the real world, however, have degree distributions differ from this. Most are highly right-skewed, meaning that a large majority of nodes have low degree, but a small number, known as "hubs", have high degree. Some networks, notably the Internet, the WWW, and some social networks are found to have degree distributions that approximately follow a power law: $P_k \sim k^{-\gamma}$, where γ is a constant. Such networks are called scale-free networks and have attracted particular attention for their structural and dynamical properties. We will discuss them in the following part.

Shortest Path is very important especially in the transport and communication within a network. In graph theory, the shortest path problem is the problem of finding a path between two vertices (or nodes) in a graph such that the sum of the weights of its contained edges is minimized. An example is finding the quickest way to go from one location to another on a road map. In this case, the vertices represent locations and the edges represent segments of road. For such a reason,

shortest paths have also played an important role in the characterization of the internal structure of a graph. It is useful to represent all the shortest path lengths of a graph G as a matrix N in which the entry d_{ij} is the length of the geodesic from node i to node j .The maximum value of d_{ij} is called the diameter of the graph.

Betweenness Centrality. Betweenness is number of edges on the shortest path from one vertex to another vertex. Then the betweenness centrality is a measure of a node's centrality in a network equal to the number of shortest paths from all vertices to all others that pass through this node. The same definition also applies to edge betweenness centrality. Betweenness centrality is a more useful measure of the load placed on the given node in the network as well as the node's importance to the network than just connectivity. The latter is only a local effect while the former is more global to the network.

Clustering, also known as transitivity, is a unique property of acquaintance networks, where two individuals with a common friend are most likely to know each other. For a usual graph G, transitivity means there exist high number of triangles. This can be quantified by defining the transitivity T of the graph as the relative number of transitive triples, i.e. the fraction of connected triples of nodes which also form triangles,

$$T = \frac{3 \times \sharp\text{of triangles in } G}{\sharp\text{of connected triples of vertices in } G}, \tag{1.3}$$

where the factor 3 in the numerator compensates for the fact that each complete triangle of three nodes contributes three connected triples, one centred on each of the three nodes, and ensures that$0 \leq T \leq 1$, with $T = 1$ for k_N.

Network Motifs are connectivity-patterns (sub-graphs) that appear much more often in real networks than they do in random networks. In biology, ecology

and other fields, most networks have been found to demonstrate a small set of network motifs. Surprisingly, the networks seem to be highly composed of these network motifs, appearing again and again. Various kinds of network seems to have its own set of typical motifs (ecological networks have different motifs than gene regulation networks, etc.). These small ones can be considered as simple building blocks from which the network is composed. This idea was first presented by Uri Alon and his group [22][23] who studied small motifs in biological and other networks. The research of the significant motifs in a graph G is based on matching algorithms counting the total number of occurrences of each n-node subgraph M in the original graph and in the randomized ones. (Fig. 1.3 [22])

Community Structure refers to the occurrence of groups of nodes in a network that are more densely connected internally than the rest of the network. Example image shown in the Fig. 1.4 [17] is the hierarchical organization displayed by most networked systems in the real world [24]. Real networks are usually composed by communities including smaller communities, which in turn include smaller communities, etc. The human body offers an example of hierarchical organization: it is composed by organs; organs are composed by tissues, tissues by cells, etc. The generation and evolution of a system organized in interrelated stable subsystems are much quicker than if the system were unstructured, because it is much easier to assemble the smallest subparts systems [25],

One of the most crucial features shown by real-world networks is the existence of modular or community structures[24]. The study of community structures helps to explain the organization of networks and eventually could be related to the functionality of groups of nodes.

Regardless of the type of real-world network in terms of the degree and other

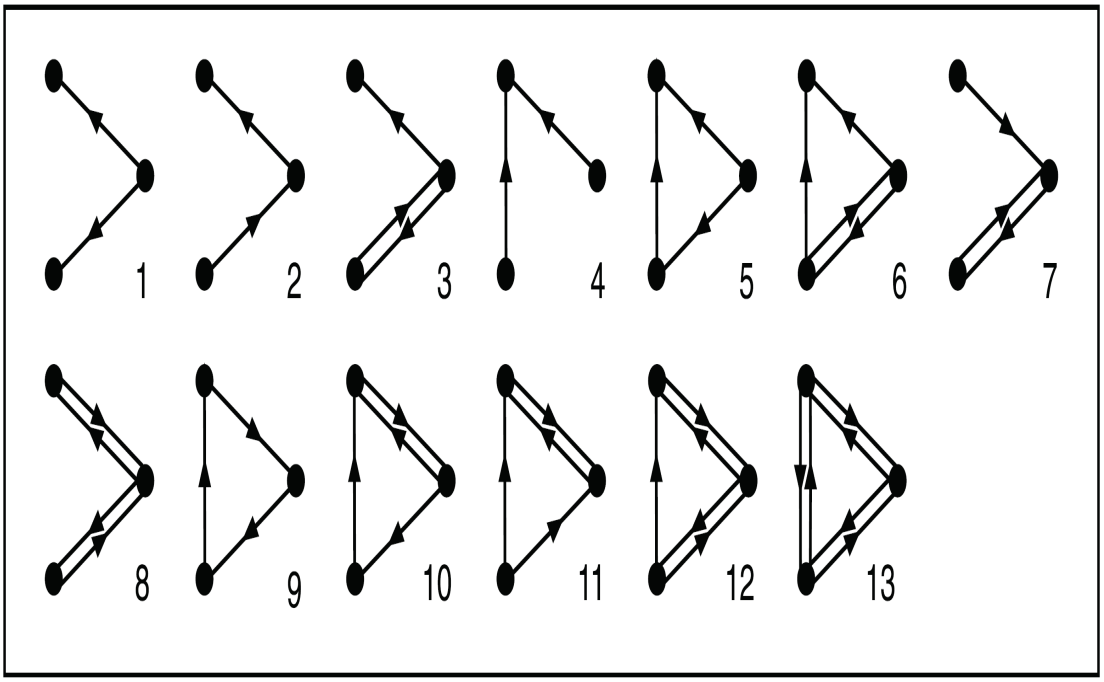

Figure 1.3: **All 13 types of three-node connected sub-graphs defined as motifs in protein-protein interaction networks in biological networks.** Adapted from Ref. [22].

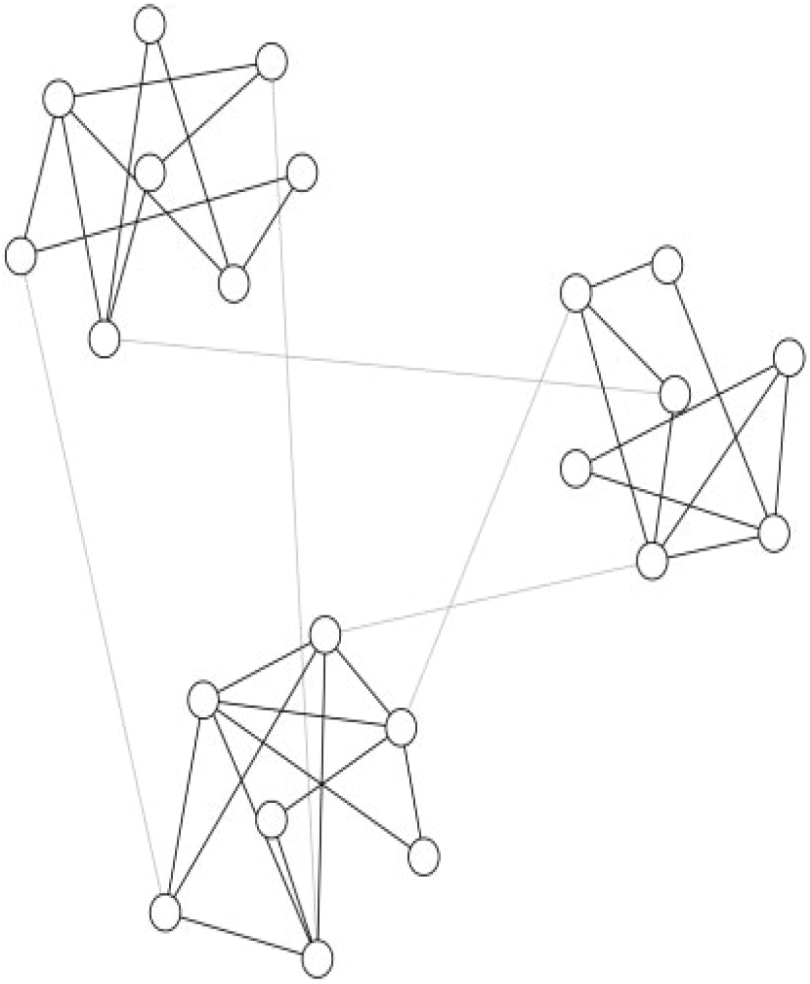

Figure 1.4: **A schematic representation of a network with community structure.** In this network there are three communities of densely connected vertices (circles with solid lines), with a much lower density of connections (gray lines) between them. Adapted from Ref. [17].

structural properties [1], it is possible to distinguish communities throughout the whole network[17]. Recently, most studies focus on biological and social networks. There are also few applications to other types of networks.

For example, the standard Girvan-Newman algorithm has been proved to be reliable to detect functional modules in biological PPIs networks [17]. Edges between modules are important points of communication. Applying the algorithm by Girvan and Newman with a modified definition of edge betweenness, Chen and Yuan [26] set up an novel functional modules research in yeast. In this work, Chen and Yuan were able to make predictions of the unknown function of some genes, based on the structural module they belong to. Gene function prediction is the most promising outcome deriving from the application of clustering techniques to PPIs. For the Social Networks, it also has been extensively studied for decades [27], such as Blondel et al. have analyzed a network of mobile phone communications between users of a Belgian phone operator [28].

1.2.2 Models of Complex Networks

A Random Graph, in mathematics, is a graph that is generated by some random process. The systematic study of random graphs was initiated by Erdös and Rényi in 1959 when they used probabilistic methods study on the properties of graphs as a function of the increasing number of random connections.

In their first article, Erdös and Rényi proposed a model to generate random graphs with N nodes and M edges, henceforth we call Erdös and Rényi (ER) random graphs. A random graph is gained by starting with a set of n vertices and adding edges between them randomly. Different random graph models produce different probability distributions on graphs. The Erdös and Rényi model denoted

by $G_{n,M}$, assigns equal probability to all graphs with exactly M edges. The model can be viewed as a snapshot at a particular time M of the random graph process, which is a stochastic process that starts with n vertices and no edges, and at each step adds one new edge chosen uniformly from the set of missing edges.

ER random graphs are the best studied among graph models, although they do not reproduce most of the properties of real networks.

A small-world network is a type of mathematical graph in which most nodes are not neighbors of each other, but most nodes can be reached from every other by a small number of hops or steps. Specifically, a small-world network is defined to be a network where the typical distance L between two randomly chosen nodes increases proportionally to the logarithm of the number of nodes N in the network [2].

The main mechanism to construct small-world networks is the Watts-Strogatz methods. The Watts and Strogatz (WS) model has both the small-world property and a high clustering coefficient [2]. The method to construct the WS model is based on a rewiring procedure of the edges by using a probability p. Starting with a N nodes regular lattice ring, in which each node is symmetrically connected to its $2m$ nearest neighbors for a total of $D = m \cdot N$ edges. Then, for every node, each link connected to a clockwise neighbor is rewired to a randomly chosen node with a probability p. Notice that for $p = 0$ we have a regular lattice, while for $p = 1$ the model produces a random graph with the constraint that each node has a minimum connectivity $k_{min} = m$. With the probability p, the procedure gave rise to graphs with the small-world property and a high clustering coefficient. Alternative procedures for constructing small-world networks, based on adding edges instead of rewiring, have also been proposed [29] [30] [31].

As observed in [2], the small-world property results from the drop in shorted path length L as soon as p is slightly larger than zero. This is because the rewiring of links creates long-range edges (shortcuts) that connect distant nodes. The effect of the rewiring procedure is highly nonlinear on L, and not only affects the nearest neighbor structure, but it also gains new shortest paths to the next-nearest neighbors and so on. On the other side, an edge redirected from a clustered neighborhood to another node has, at most, a linear effect on C. That is, the transition from a linear to a logarithmic behavior in $L(p)$ is faster than the one with the clustering coefficient $C(p)$. This leads to the appearance of a region of small (but non-zero) values of p, where one has both small path lengths and high clustering.

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. A large amount of works on the characterization of the topological properties of real networks, such as World Wide Web links, biological networks, and social networks, has stimulated the need to construct graphs with power law degree distributions.

The fraction $P(k)$ of nodes in the network having k links to other nodes prefer to attach large values of k as

$$P(k) \sim ck^{-\gamma}, \tag{1.4}$$

c is a normalization constant. The value of γ is typically in the range $2 < r < 3$, although occasionally it may be out of these bounds.

It is quite easy to get a graph with a power-law degree distribution which can be treated as a special case of the random graphs with a given degree distribution. Dangalchev (2004) gives examples of generating static scale-free networks. Another possibility (Caldarelli et al. 2002) is to consider the structure as static and draw a

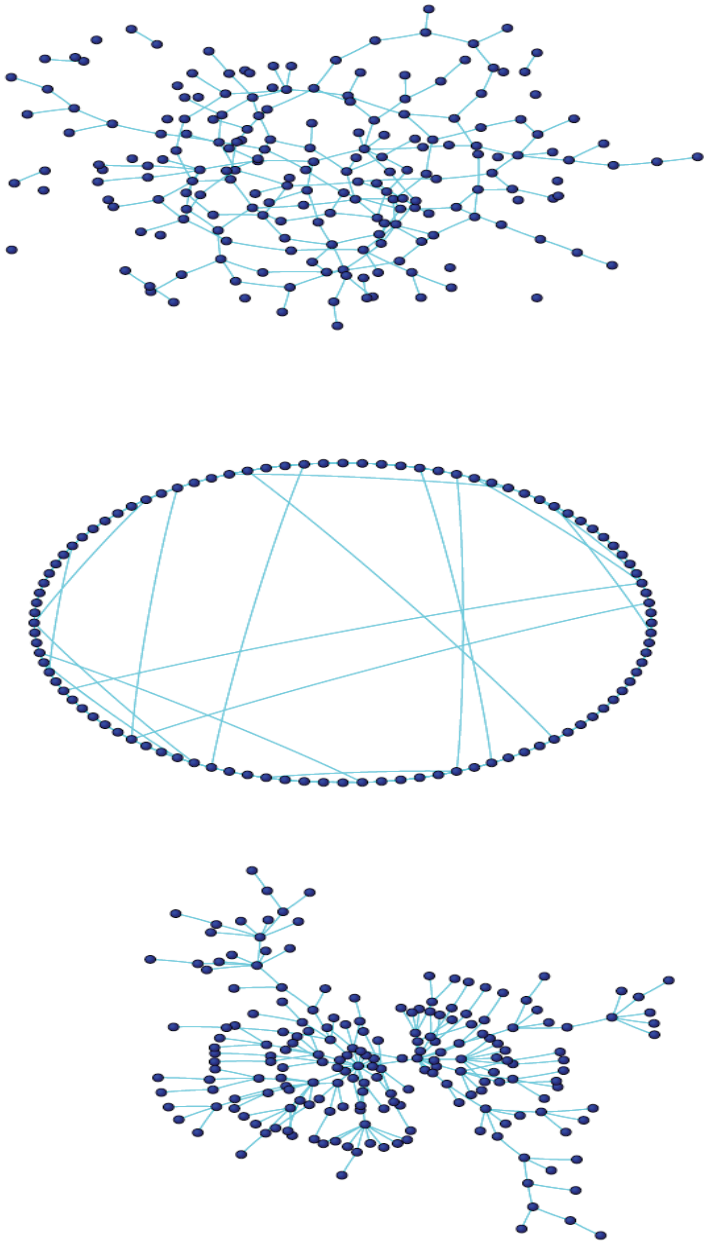

Figure 1.5: **Basic models of complex networks.** . (Top) Erdös-Rényi random graph with 100 vertices and a link probability $p = 0.02$. (Center) Small world graph Watts-Strogatz, with 100 vertices and a rewiring probability p = 0.1. (Bottom) Barabási-Albert scale-free network, with 100 vertices and an average degree of 24 . Courtesy by J. J. Ramasco [37]. Adapted from Ref. [24].

link between vertices according to a particular property of the two vertices involved. Once the statistical distribution for these vertices properties (fitnesses) is given, it turns out that static networks can develop scale-free properties. We define such graphs as static scale-free to distinguish them from models of evolving graphs.

Static scale-free graphs are good models for all cases in which growth or aging processes do not play a dominant role in determining the structural properties of the network. However, there are many examples of real networks in which the structural changes are ruled by the dynamical evolution of the system. Usually, preferential attachment and the fitness model have been proposed as mechanisms to explain conjectured power law degree distributions in real networks.

In 1999, Barabasi and Albert mapped the topology of a portion of the World Wide Web [3], finding that some nodes, which they called "hubs", had many more edges than others and that the network as a whole had a power-law distribution. After finding that a few other networks, including some social and biological networks, also had heavy-tailed degree distributions, Barabasi and Albert coined the term "Scale-Free Network" to describe the class of networks that exhibit a power-law degree distribution.

Then, Barabási and Albert proposed a generative mechanism model to explain the appearance of power-law distributions which based on two basic ingredients: growth and preferential attachment. This is essentially the same as that proposed by Price in 1976 [32] to explain the power laws. The same author also mentioned this in 1965, one decade earlier, in citation networks (both for the in-degree and the out-degree distributions) [33].

The most notable characteristic in a scale-free network is the highest-degree nodes which are often called "hubs", and are thought to serve specific purposes in

their networks, although this depends greatly on the domain.

The scale-free property strongly correlates with the network's robustness to failure. It turns out that the major hubs are closely followed by smaller ones. These ones, in turn, are followed by other nodes with an even smaller degree and so on. This hierarchy allows for a fault tolerant behavior. If failures occur at random and the vast majority of nodes are those with small degree, the likelihood that a hub would be affected is almost negligible. Even if a hub-failure occurs, the network will generally not lose its connectedness, due to the remaining hubs. On the other hand, if we choose a few major hubs and take them out of the network, the network is turned into a set of rather isolated graphs. Thus, hubs are both a strength and a weakness of scale-free networks. These properties have been studied analytically with percolation theory by Cohen et al[34] [35] and by Callaway et al.[36].

Another important feature of scale-free networks is the clustering coefficient distribution, which decreases as the node degree increases. This distribution also follows a power law.(Network models see Fig. 1.5 from Ref. [24])

1.3 Dynamic-based Structure Measures of Complex Network

In the last two sections, we have introduced the basic concepts and models of complex networks. We realize that the structural measures of complex networks are the cornerstone to understand the relations between the structures, dynamics and functions. So a very important thing to do the research of complex networks is to find a systematic measures for characterizing network structures.

Although great progresses have been achieved, the measures of complex networks are not yet fully understood. Furthermore, the measures of network structures, such as the degree distribution, cluster coefficient, motif, and the patterns at different scales, are generally a simple application of the concepts in graph theory, bioinformatics, social science and fractal theory, namely, they are static, not dynamic-based. We cannot expect simple and reasonable relations between the structure measures, functions and the dynamical processes on networks.

Dynamics on complex networks, as the bridge between structures and functions, can be regarded as the transport processes of mass, energy, signal and/or information at different structure scales [7, 8]. In the progress of dynamic transport, the structure determine the function of the networks, meanwhile, the realizations of the functions depend on the progress of the dynamics. So the lack of powerful tools to characterize network structures is an essential bottleneck to understand dynamical processes on networks. One typical example is the synchronizabilities of complex networks. Detailed works show that almost all the structure measures affect the synchronizabilities in complicated ways [11], based upon which we cannot reach a clear picture of the mechanisms for synchronization processes on networks.

The structures of complex networks can induce nontrivial properties to the physical processes occurring on them. The physical processes in turn can be used as the probe to capture the structure properties. Well studied dynamical processes, such as the diffusive process [38], random walks [12, 13] and the Boolean dynamics [14], can be good candidates as probes. To cite an example, the random walks on complex networks that biased towards a target node show a localization-delocalization transition[12].

Random Matrix Theory (RMT) was initially developed to understand

the statistical properties of nuclear spectra. It made successful predictions for the spectral properties of different complex systems such as disordered systems, quantum chaotic systems, large complex atoms, quantum graphs etc., in the last few decades [15] [39].

Recently, the RMT theory has been proposed to capture the structure and dynamical properties of complex networks, such as: price fluctuations in data of stock market [40], human Brain EEG data [41], epidemic disease data, variation of atmospheric parameters [42], and many others.

In the present thesis, we study the properties of complex networks by means of their eigenvalue spectra and especially their localized eigenvectors. Based on that, we are able to construct the dynamics-based structure measures of complex networks. We also analyze very successfully evolving and directed networks.

1.3.1 Random Matrix Analysis of Complex Networks

As we have discussed before, tremendous activities have been put to the network studies. Apart from the above mentioned studies which focus on direct measurements of structural properties of networks, such as degree distribution, cluster coefficient, motif, and communities, moreover, there exists a vast literature proving that properties of networks or graphs could be well characterized by the spectrum of associated adjacency A and Laplacian L matrix [43].

In the present thesis, we map a complex network of N coupled identical oscillators to an artificial molecule: the nodes as atoms and the edges as the bonds between them. The topological structure of the molecule can be described by an adjacency matrix A or a Laplacian matrix L. We consider an electron moving on it, and after several steps simplification(described in Chapter 2), then the system'

tight-binding Hamiltonian could be written in $H = A(orL)$.

For an adjacency matrix A of a network, $A_{ij} = 1$ if i and j nodes are connected and 0 otherwise. Laplacian of network has been defined in various ways (depending upon the normalization) in the literature. Here we define the Laplacian matrix as: the off-diagonal elements of L are $L_{i,j} = L_{j,i} = -1(0)$ if the nodes i and j are connected (disconnected), respectively; The diagonal elements are $L_{ii} = -\sum_{j \neq i} L_{ij} = k_i$, where k_i is the number of the nodes connected directly with the node i (node degree).

For an undirected network, adjacency and Laplacian both are symmetric matrices and consequently have real eigenvalues. Eigenvalues of the networks are called graph spectra and they give information about some basic topological properties of underlying networks [43]. The spectral analysis of the networks shows rich information about the topological structure and diffusion-localization processes.

During the last decades, several important applications of the spectral analysis theory in physics and chemistry problems have been discovered. For example liquid flowing through several connected pipes are expressed by a serial of linear differential equations. The underlying graph can be expressed by the corresponding Laplacian matrix. Also the speed of the liquid convergence flowing process towards an equilibrium state could be measured by the second largest eigenvalue of graph Laplacian [44].

The second largest eigenvalue of graph Laplacian is also called the algebraic connectivity of a graph and is used to understand behavior of dynamical processes on the underlying networks [45][46][47]. Moreover, the spectra of the Laplacian matrix of networks have also been studied greatly to understand synchronization of coupled dynamics on networks [45][47]. For example recently eigenvalues of the

Laplacian have been shown to have high influence on the synchronizability of the network [48]. Also, system analysis of eigenvalues, particularly at 0 and 1, have direct relations with the properties of graphs [49, 50]. Studies of spectral properties of the complex networks may also have a general theoretical interest.

As for different model networks and real world networks, the properties of the networks could also be well characterized by the spectral density of adjacency matrix and Laplacian L matrix. For a random network, whose elements are randomly 0 or 1, the spectral density follows the semicircular law [51]. Interestingly, this result matches quite well with a distinguished result in Random Matrix theory (RMT) about the spectral density of a random matrix, whose elements are Gaussian distributed random numbers, following Wigner's semicircular law [15] . With the increasing availability of large maps of real world networks, the analysis of spectral densities of real world networks and model networks having real-world properties have also begun [8] [52] [53] [54] [55] [56] [57] [58]. These analysis show that the spectral densities of model networks and real world networks are not semi-circular, instead they have some specific features depending on the details of the networks. For example, small-world model networks show very complex spectral densities with many sharp peaks, while spectral densities of scale-free model networks exhibit triangular distribution [51].

One of the main advantages of RMT approach is that depending on the nature of eigenvalues correlations, we can separate system dependent part from random universal part, which is interconnected due to the complexity of the system. RMT analysis for the various networks shows that the Nearest Neighbor Level Spacing (NNLS) of complex networks also follow universal GOE statistics of RMT [53] [57] [59]. This finding suggests that different results of GOE statistics, which have

successfully been applied to understand the systems coming from various fields starting from nuclei to the stock-market, can be applied to study networks as well. The NNLS distribution is also used to identify collective chaos in the corresponding classical dynamics on the complex networks [60].

Nearest Neighbor Level Spacing

One of the most important quantities in the Random Matrix Theory (RMT) is the Probability Distribution Function (PDF) for the nearest neighbor level spacing (NNLS) of the spectrum. The nearest neighbor level spacing distribution(NNLSD) $p(s)$ is the most often used to study the short-range fluctuations in the spectrum. The NNLSD gives probability for finding neighboring eigenvalues with a given spacing, which is equal to the probability density for two neighboring levels ξ_n and ξ_{n+1} that have the spacing s.

In order to get universal properties of the fluctuations of the eigenvalues, it is customary in RMT to unfold the eigenvalues by a transformation **Unfolding Procedure**.

A measurement gives an ordered sequence of energies $(E_1, E_2, \cdots, E_N)$ form the stick spectrum or spectral function [39]. Then we define the cumulative spectral function,

$$\eta(E) = \int_{-\infty}^{E} S(E')dE' = \sum_{n=1}^{N} \Theta(E - E_n). \tag{1.5}$$

The number of the energy levels less than or equal to E is counted in this function. And the function is also regarded as the staircase function. It is break up into a smooth part $\xi(E)$ and a fluctuating part $\eta_f(E)$,

$$\eta(E) = \xi(E) + \eta_f(E). \tag{1.6}$$

The smooth part $\xi(E)$ is given by the cumulative mean level density,

$$\xi(E) = \int_{-\infty}^{E} R_1(E')dE'. \tag{1.7}$$

Fig. 1.6 (from reference [39]) shows an example for an experimentally obtained staircase function and its smooth part. To unfold the spectrum, the sequence $(E_1, E_2, \cdots, E_N)$ is mapped onto the numbers $(\xi_1, \xi_2, \cdots, \xi_N)$ with

$$\xi_n = \xi(E_n), n = 1, \cdots, N, \tag{1.8}$$

the cumulative spectral function in these new variables reads,

$$\eta(\xi) = \xi + \eta_f(\xi). \tag{1.9}$$

Using the unfolded spectra, we calculate the spacing as,

$$s = \xi(E_{n+1}) - \xi(E_n), n = 1, \cdots, N, \tag{1.10}$$

The unfolded spectrum's mean level density, i.e. the derivative of the smooth part with respect to ξ, is unity, as required. Henceforth, measurement and theoretical prediction could be directly compared with each other.

Then for the NNLSD function $p(s)$ and its first moment are normalized to unity,

$$\int_o^{\infty} p(s)ds = 1, \int_o^{\infty} sp(s)ds = 1, \tag{1.11}$$

and it follows two universal properties depending upon underlying correlations among the eigenvalues. For the correlated eigenvalues, NNLSD follows Gaussian

orthogonal ensemble (GOE) statistics of RMT, whereas for the uncorrelated eigenvalues it follows Poissonian statistics. Generally, for an intermediate state, the PDF obeys the Brody distribution

$$U(s) = \xi\beta s^{\beta-1} exp[-\xi s^{\beta}], \tag{1.12}$$

where s is the NNLS and $\xi = [\Gamma(\frac{\beta+1}{\beta})]^{\beta}$ is the characteristic distribution width. The Poisson and the Wigner-Dyson distributions are the two extremes with $\beta = 1$ and $\beta = 2$, respectively.

Spectral Rigidity: Long-range spectral observable

NNLSD carries information only for the local correlations among the eigencalues, but do not tell about the correlation between two far away ones. Therefore, even though NNLSD follows GOE statistics of RMT, other properties may show deviations, which suggests that one cannot rely on NNLSD results exclusively.

To probe long-range correlations as well, researchers usually also consider the spectral rigidity, measured by the well known $\triangle_3$-statistic of RMT, which is more sensitive for RMT properties of the matrix under investigation [15] [59] , and more suitable to analyze larger mixing, in general, in the case for the real world networks.

The $\triangle_3$-statistic measures the least-square deviation of the spectral staircase function representing the cumulative density $N(\bar{\lambda})$from the best straight line fitting for a finite interval L of the spectrum.

$$\triangle_3(L; x) = \frac{1}{L} \min_{a,b} \int_x^{x+L} [N(\bar{\lambda} - a\bar{\lambda} - b]^2 d\bar{\lambda}), \tag{1.13}$$

where a and b are obtained from a least-square fit. Average over several choices of x gives the spectral rigidity $\triangle_3(L)$. For GOE, $\triangle_3(L)$ depends logarithmically on L,

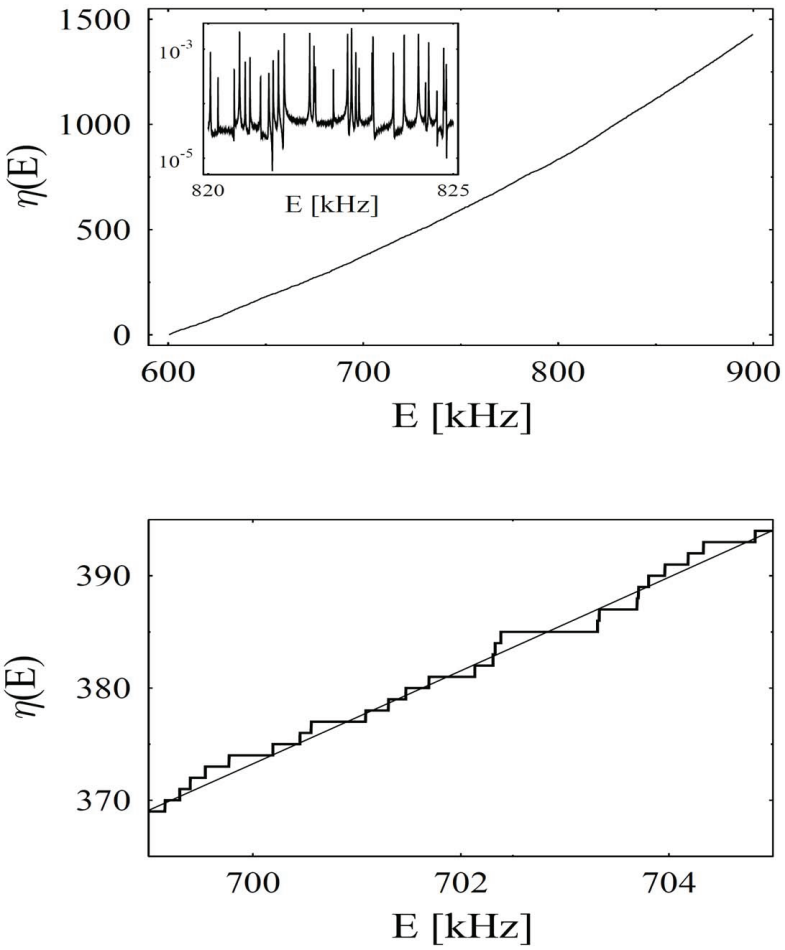

Figure 1.6: **Example of an experimentally obtained staircase function.** The top figure shows the cumulative spectral function (E) for a spectrum of 1428 elastomechanical eigenfrequencies of a resonating quartz block in the frequency range between 600 kHz and 900 kHz. The frequency is denoted by E. Due to the high number of levels, the staircase function appears as a smooth line. The smooth part $\xi(E)$ is a polynomial whose coefficients were found by a fit. The corresponding section of $\xi(E)$, i.e. the polynomial fit, is drawn as a thin line. We notice that $\xi(E)$ is obtained by fitting to the entire cumulative spectral function, not only to this section. Picture courtesy of Ref. [61]. Adapted from Ref. [39].

$$\triangle_3(L) \cong \frac{1}{\pi^2} \ln L. \tag{1.14}$$

It is found that the spectral rigidity of the complex networks follows RMT prediction, with scale depending upon the properties of the networks.

1.3.2 Evolution of Complex Networks

Evolving networks are networks whose structures evolve with time. They are natural extension of network science, since most real networks are dynamic. To detect the evolutionary history of a network is of a great importance. We will present the spectral-analysis based method to address this fundamental question in networks in Chapter 3.

For the evolving networks, the changes of topology can occur if some nodes (edges) appear (disappear), and the types or weights of nodes and edges can also change even if the topology stays static. Usually all of these processes occur simultaneously, such as in epidemic networks where people infected by or recovered from disease over time, thereby getting or losing edges, and some new infected people become part of the new epidemic network, changing the nodes in the network. Or in social networks where people collaborate or stop collaboration with others over time, thereby creating and destroying edges, and some people become part of new social networks or leave their networks, changing the structure of the network. Evolving network concepts build on established network theory and are now being introduced into studying networks in many diverse fields [62].

The Barabasi-Albert (BA) model is the first model to derive the network topology from the way the network was generated with nodes and edges being added over time. Such as new web pages are added over time, and each new page

is more likely to link to highly visible hubs like Google which have high degree distributions than to nodes with only a few links. However, the model makes only the simplest assumptions necessary for a scale-free network to emerge, say there is only linear growth and linear preferential attachment. This simple model does not capture variations in the shape of the degree distribution, or the size independent clustering coefficient. So the original model has since been modified to capture the properties of evolving networks by introducing a few new properties.

The most common way to view evolving networks is by considering them as successive static networks. This could be viewed as the composing a motion picture. Many simple parameters exist to describe a static network (number of nodes, edges, path length, connected components), or to describe unique nodes in the graph such as the number of links or the clustering coefficient. These properties can then individually be researched as a time series using signal processing notions[63]. Unfortunately, the analogy of snapshots to a motion picture also reveals the main difficulty with this approach: the time steps employed are very rarely suggested by the network and are instead arbitrary. Using extremely small time steps between each snapshot preserves resolution, but may actually obscure wider trends which only become visible over longer timescales. Conversely, using larger timescales loses the temporal order of events within each snapshot. Therefore, it may be difficult to find the appropriate timescale for dividing the evolution of a network into static snapshots.

Using successive snapshots is that only slight changes in network topology can have large effects on the outcome of algorithms designed to find different groups of nodes existed from different amount of time. Therefore, it is necessary to find a new methods which permits following the evolution of the community through a

set of rules such as birth, death, merge, split, growth, and contraction[64].

Almost all real world networks are evolving networks since they are constructed over time. By varying the respective probabilities described above, it is possible to use the expanded BA model to construct a network with nearly identical properties as many observed networks. Moreover, the concept of scale free networks shows us that time evolution is a necessary part of understanding the network's properties, and that it is difficult to model an existing network as having been created instantaneously. Real evolving networks which are currently being studied include social networks, communications networks, the internet, the movie actor network, the world wide web, transportation networks, Metabolic, protein protein interaction networks, and genetic networks. There also include spreading evolving networks such as epidemic, rumor, and a piece of joke spreading networks.

1.4 Thesis Outline

The subsequent chapters in this dissertation are organized as follows:

In Chapter 2, firstly, for undirected networks, we shall study the structural characteristics of complex networks using the representative eigenvectors of the adjacent matrix. The probability distribution functions of the components of the representative eigenvectors are proposed to describe the localization on networks where the Euclidean distance is invalid. Several quantities are used to describe the localization properties of the representative states, such as the participation ratio, the structural entropy, and the probability distribution function of the nearest neighbor level spacings for spectra of complex networks. Whole-cell networks in the real world and the Watts-Strogatz small-world and Barabasi-Albert scale-free

networks are considered.

Then for the directed networks with inhibitory and excitatory couplings spectra analysis, we shall investigate in particular eigenvector localization properties of random networks for different values of correlation among their entries. Spectra of random networks with completely uncorrelated entries show a circular distribution with delocalized eigenvectors, whereas networks with correlated entries have localized eigenvectors. In order to understand the origin of localization, we track the spectra as a function of connection probability and directionality. As connections are made directed, eigenstates start occurring in complex-conjugate pairs and the eigenvalue distribution combined with the localization measure shows a rich pattern.

The networks have nontrivial localization properties due to the nontrivial topological structures. It is also found that the ascending-order-ranked series of the occurrence probabilities at the nodes behave generally multi-fractal. This characteristic can be used as a structural measure of complex networks.

In Chapter 3, we present a very interesting method about how to detect the evolutionary history of a network with spectral analysis. This is considered as a new application of the spectra. In particular, we find that there are complex networks in the real-world for which there is a positive correlation between the eigenvalue magnitude and node age. In situations where the network topology is unknown but short time series measured from nodes are available, we suggest to uncover the network topology at the present (or any given time of interest) by using compressive sensing and then perform the spectral analysis.

Finally, a general summary of this dissertation and a perspective of my PhD research project are given in **Chapter 4**.

Chapter 2

Localizations on Complex Networks

In this chapter, we shall study the structural characteristics of un-directed complex networks using the representative eigenvectors of the adjacent matrix and also apply the spectra analysis for directed complex networks with inhibitory and excitatory coupling as well.

This Chapter is based on my two published papers [56] [57]. The first part is from paper [56], whereas the second part is from part of the paper [57].

In the first part of this chapter, we introduce the concept of localization on complex networks for the un-directed networks. The occurrence probabilities on the nodes are proposed to describe quantitatively the localization effects. First, the methods to measure the localization properties are described in details. The participation ratio, the structural entropy, and the probability distribution function of the nearest neighbor level spacing of spectra are used to illustrate the localization in a global way. Then, wavelet transform is used to find the detailed structural

properties of the probability distribution function of the occurrence probabilities on the nodes. Followed as examples, we consider the Watts Strogatz small-world, the Barabasi-Albert scale-free model networks and whole cell networks in the real world. Finally the results are shown in the last part. We will show that the global symmetries in networks can induce multi-fractal structures in the eigenvectors. As a conclusion, the nontrivial structures of complex networks can induce significant localization, which in turn can be used as a global measure of the structural symmetries.

As for the directed networks in the second part, we will evaluate the applicability of RMT for distribution of eigenvalues of directed networks where entries in corresponding matrix take values motivated by inhibitory and excitatory coupling between nodes. We study in particular eigenvector localization properties of various model networks for different values of correlation among their entries. Spectra of random networks with completely uncorrelated entries show a circular distribution with delocalized eigenvectors, whereas networks with correlated entries have localized eigenvectors. The main goal of this part of work is to investigate the changes in spectral properties of corresponding matrix as underlying directed network deviates from the perfect community structure.

2.1 Localizations on Undirected Complex Network

Impacts of network structure on the dynamical processes attracts special attentions in recent years [65], such as the synchronization of dynamical systems on networks, the epidemic spreading on networks, and the PPI interactions on

networks. In the mean time, different models have been proposed to study and understand networks which have underpinning network structures. Watts and Strogatz proposed the 'small-world network' [2], which captures randomness and uncovers the larger clustering of real-world networks. Barabasi and Albert generate a model to capture degree distributions of real world networks [3]. According to this model, there exist hubs on the networks. Since then tremendous studies about structure and dynamics of networks came out. Furthermore, there exists extensive literature demonstrating that the properties of networks are well characterized by the spectrum of associated adjacency matrices [8][11][52][53][54][55][56][57][58].

In the first part of the present chapter, we map networks to a large molecular clusters, namely, the nodes and edges to atoms and bonds between them, respectively. The localization properties of electrons in the cluster can be used as measures of the structure properties of the networks. We try to detect the global symmetries from the spectra and the eigenvectors of complex networks. Very recently, much attentions have been focused on detecting global characteristics embedded in spectra of complex networks due to their potential application in understanding the organization mechanisms and the synchronization dynamics of complex networks. To our best knowledge, it is the first time to detect the global characteristics of complex networks from the eigenvectors which contain more information about the system than eigenvalues.

Besides as a measure of network structures, the structure-induced localization may have potential application in understanding the electronic properties of materials such as conductive polymers and carbon nanonets. The intra-chain windings in conductive polymers can introduce long-range edges into the original one dimensional (1-D) systems, resulting in nontrivial network structures [66, 67]. It is

also found that random networks of carbon nanotubes, called nanonets, can mimic a variety of basic electronic functions from the conductive properties of metals to the less conductive characteristics of semiconductors [68]. Indeed, nanonets have paved the way for the carbon to serve as the foundation for future electronic devices. The effect of network structures on electronic properties is one of the most active topics in recent years [7, 69].

2.1.1 Methods

We consider a undirected complex network with N identical nodes, whose topological structure can be described by an adjacent matrix A. The elements A_{ij} are 1/0 if the nodes i and j are connected/disconnected, respectively. If we consider the nodes as atoms and the edges as bonds, the network can be mapped to a large molecule [58]. For an electron moving in such a molecule, the tight-binding Hamiltonian is,

$$\mathcal{H} = \sum_{n=1}^{N} \varepsilon_n \cdot |n\rangle \langle n| + \sum_{m \neq n}^{N} A_{mn} \cdot t_{mn} \cdot |m\rangle \langle n|, \tag{2.1}$$

where ε_n is the site energy and $t_{mn} \cdot A_{mn}$ the hoping integral for the bond between sites m and n.

A tight-binding Hamiltonian of Eq. 2.1 is usually used to study the disorder-induced localizations. In the present form, the matrix A is explicitly introduced to describe the structure of the system. For a one dimensional (1-D) perfect regular lattice, we have $\varepsilon_n = \varepsilon, t_{mn} = t$ and $A_{mn} = \delta(m - n \pm 1)$. The Bloch wave function of an electron extends all over this perfect regular lattice. Disorder structures can induce a transition from extended to localized states. The wave function

for a localized state decreases exponentially with the distance from its center. The disorder effects include the random distributions of the site energies (ε_n), the hoping integrals (t_{mn}) and the edges in structures (A_{mn}). The disorders come from the different kinds of atoms on the lattice points, the differences of the separations of successive lattice points and the randomness in structures. At the same time, there may be some symmetries in the distributions of the site energies, the hoping integrals and the edges, which may lead to delocalization of the wave functions.

In the usual Anderson model [70], the disorder effect due to the random distribution of the site energies is considered, i.e., ε_n is a random variable satisfying a certain probability distribution function while $t_{mn} = t, A_{mn} = \delta(m - n \pm 1)$. The site energies may obey a special distribution rather than that in the Anderson model, as a periodic [71] or a power-law [72] function. In literatures [73, 74], a one-dimensional quasi-crystal model is introduced that the separation of two successive lattice points takes one of the two values u and v. This model considers the disorder effect of the distribution of the hoping integrals. We have ε_n=const., $t_{mn} = t_u$ or t_v and $A_{mn} = \delta(m - n \pm 1)$. t_u and t_v are the hoping integrals corresponding to the separations u and v, respectively. It is found that quantum systems with quasi-periodic structures will be in an intermediate state, which can be described with critical wave functions. A critical wave function obeys a power-law with respect to the distance from its center.

To investigate the problems as vibration spectra of glasses, instantaneous normal modes in liquids, electron hopping in amorphous semiconductors and combinatorial optimization, Euclidean random matrix (ERM) models are widely used in literatures [75], in which the disorder is due to the random positions of the sites, and the matrix elements are given by a deterministic function of the distances.

The models mentioned above generally focus on the disorder effects of the site energies and the hopping integrals. These models have also been extended to nontrivial structured systems such as the Cayley tree [76] and the small-world networks [66]. Nontrivial effects of the structures of the systems are reported, but the interplay between the disorders due to the site energies and that due to the structures makes it difficult to distinguish the structure disorder effect from the site energy disorder effect.

In the networks considered in this part, however, the nodes are all identical and the disorder effect comes from the nontrivial topological structure. We focus our attentions on the disorder effect of the network structure, that is, we assign $\varepsilon_n = 0$ and $t_{mn} = 1$, which leads to $H = A$. The localization on the network refers to the network structure-induced characteristics of the wave functions for this system. The usual Anderson model [70] is a special case that in the networks there exist connections only between the nearest neighbors in Euclidean space.

Statistically, the structures of networks should display certain symmetries due to the general rules obeyed in the construction of the networks. Recent works demonstrate that many theoretical and real world networks have statistically self-similar structures [6]. Therefore, there are two competitive mechanisms determining the wave function property, the randomness of the bonds in the networks tends to cause localization of wave function, whereas the symmetries of the networks intend to make wave function extended. We thus expect rich structures embedded in the wave functions. As it is well known, aperiodic crystals lead to the fractal wave functions [74, 77]. An interesting question is then, how the global symmetries of networks affect the localization properties. The localization can be used as a probe of the characteristics of the network structures.

The states in the center of the energy band have the best chance to remain as extended for a moderately disordered system. The eigenvector corresponding to the special eigenvalue close to the center of the spectrum for a network, denoted by E_c, is employed as the representative state to illustrate the characteristics of the considered system.

In the traditional study of wave function localization, the physical systems have deterministic structures in real world Euclidean space, which leads to natural definitions of the localized, intermediate and extended states of the systems. Obviously, these definitions are invalid for general complex networks without deterministic structures in Euclidean space. In this thesis, we describe the localization effects with the probability distribution function (PDF) of the occurring probabilities at the nodes, i.e., the values of the components for the representative eigenvector. Based on the PDF of the occurring probabilities, the traditional definitions are extended to a much more general version to describe the localization properties on complex networks.

In Euclidean space, for a state $\Psi(r)$, the occurring probability is $\rho(r) = |\Psi(r)|^2 \equiv F(r)$. Because the value of the distance r distributes homogeneously in the considered region, we can regard it as a homogenously distributed random variable. The direct sampling method in Monte Carlo simulations tells us that the probability distribution of ρ should be $P(\rho) \propto \frac{dF^{-1}(\rho)}{d\rho}$. Hence, it is reasonable to define the localized, critical and perfectly extended states on complex networks with the PDFs of the occurring probabilities,

$$P(\rho) \propto \begin{cases} \delta(\rho - \rho_0), & extended \\ \rho^{-(1+\eta)}\,|_{\eta>0}\,, & critical \\ \rho^{-(1+\eta)}\,|_{\eta=0}\,, & localized \end{cases} . \tag{2.2}$$

The PDF of the representative function is a very powerful measure to capture the localization properties. It can be used to find the localization properties without using distance in real world Euclidean space.

Because no derivative exists for a fractal wave-function in Euclidean space, the extension procedure in defining critical and localized states on networks can not be simply used to define fractal property on networks with the PDF of the occurring probabilities. Here, we detect directly the fractal characteristics in the ascend-ranked series of the the occurring probabilities, as described in detail in Section 2.1.4.

2.1.2 Structural Entropy

We denote the representative state with $V = (V_1, V_2, \cdots, V_N)$. The occurring probabilities at the nodes are $\rho_m = |V_m|^2, m = 1, 2, \cdots, N$. The localization extent of the state can be described quantitatively with the participation ratio [78, 79],

$$Q = \frac{1}{N \cdot \sum_{m=1}^{N} {\rho_m}^2}. \tag{2.3}$$

For a perfect extended state we have $Q = 1$, while for a state strongly localized on one node it tends to $\frac{1}{N}$. Generally, Q should be in the range of $\left[\frac{1}{N}, 1\right]$.

However, this participation ratio can capture only the primary-level complexity in the localization properties, namely, the extension of the representative eigenvector to $N \cdot Q$ nodes on the network. Many PDFs with different localization

behaviors may result in the same Q. The simplest one is a step-like function that on $N \cdot Q$ nodes the occurring probabilities are $\frac{1}{N \cdot Q}$, while on the left $N \cdot (1 - Q)$ nodes the occurring probabilities are 0.

The secondary-level complexity in the localization properties is the deviation of the PDF from the step-like function. This deviation corresponds to the shape of the PDF, which can be extracted by using the structural entropy [80],

$$S_{str} = -\sum_{m=1}^{N} \rho_m \ln \rho_m - \ln(Q \cdot N). \tag{2.4}$$

For the simple step-like condition, we have $S_{str} = 0$. $S_{str} \neq 0$ tells us the shape deviation of PDF from the simple one.

The pair of localization quantities, (Q, S_{str}), is widely used up to date to describe the localization in disordered and aperiodic systems, and the statistical analysis of spectra in diverse fields such as quantum chemistry, condensed matter physics, and quantum chaos [81, 82].

2.1.3 Statistical Properties of the Spectra

The localization property can also be described with the random matrix theory (RMT) [39, 79, 83]. RMT is initially developed to understand the energy levels of complex nuclei and other kinds of complex quantum systems. Recently, the RMT theory has been proposed to capture the structure and dynamical properties of complex networks [84].

One of the most important quantities in the theory is the PDF for the nearest neighbor level spacings (NNLS) of the spectrum. It is theoretically and numerically confirmed that at the localization and the extended states the PDFs of the NNLS should be Poisson and Wigner-Dyson distribution, respectively [85–87]. Generally,

for an intermediate state, the PDF obeys the Brody distribution,

$$U(s) = \xi\beta s^{\beta-1} exp[-\xi s^{\beta}], \tag{2.5}$$

where s is the NNLS and $\xi = [\Gamma(\frac{\beta+1}{\beta})]^{\beta}$ is the characteristic distribution width. The the Poisson and the Wigner-Dyson distributions are the two extremes with $\beta = 1$ and $\beta = 2$, respectively.

Introducing the accumulated function, $C(s) = \int_0^s U(x)dx$, some trivial calculations lead to,

$$\ln R(s) \equiv \ln\left[\ln\left(\frac{1}{1-C(s)}\right)\right] = \beta \ln s - \beta \ln \frac{1}{\xi}. \tag{2.6}$$

From this formula we can testify the Brody distribution and determine reliably the values of the parameters β and ξ.

In order to get universal properties of the fluctuation of the eigenvalues, to make the spacings s in units of local mean level spacing, we should conduct a standard procedure, called unfolding. Denoting the spectrum of a network with $\lambda_1, \lambda_2, \cdots, \lambda_N$, the accumulation density function for the spectrum is $G(\lambda_m) = m, m = 1, 2, \cdots, N$. Fitting this relation with a polynomial function, we can split it into a smooth part, G_{av}, where is the averaged integrated eigenvalue density and a fluctuation part as,

$$G(\lambda_m) = G_{av}(\lambda_m) + G_f(\lambda_m). \tag{2.7}$$

The NNLS can be obtained as,

$$s_i = G_{av}(\lambda_{i+1}) - G_{av}(\lambda_i), i = 1, 2, \cdots, N-1, \tag{2.8}$$

For a complex network, we generally have not enough knowledge on its spectrum. In the standard random theory, a random matrix has generally some perfect characteristics, e.g., all the elements are uncorrelated and obey a same isolated identical distribution.

But an adjacency matrix of a network deviates generally from the standard random matrix, because the network has non-trivial patterns. Hence, we have less knowledge on the characteristics of the spectrum. A proper way to conduct unfolding is to fit the integrated spectral distribution function with a polynomial function.

There appears a technical problem that how to select the order of the polynomial function. A proper criterion is that the order should be large enough so that the behaviors for nearest neighbor spacing distribution function keep unchanged in a considerable wide range of order. Here, the order of the polynomial function is 17. 17th order turns out to be enough for our purpose, and it is still a reasonable calculating order for the computer to do the polynomial. After the unfolding, the average level spacing will be unity, which is independent of the system.

2.1.4 Wavelet Transform

The detailed properties for the PDF of the occurring probabilities can be used as the measure of the global structure symmetries. However, determining this PDF is a nontrivial task [88]. Assume the probability values have been sorted in ascending order, namely, $\rho = \{\rho_1 \leq \rho_2 \cdots \leq \rho_N\}$, which can be regarded as the profile of the nearest spacing series, $\Delta\rho = \{\rho_2 - \rho_1, \rho_3 - \rho_2, \cdots, \rho_N - \rho_{N-1}\}$. The local structures of $\Delta\rho$ can tell us the probability distribution function of ρ. It is found that the series ρ generally behaves multi-fractal.

The wavelet transform (WT) [89] is used to detect the fractal properties embedded in the ascend-ranked series ρ. The increasing trend in the series ρ makes the box-counting-based techniques invalid to quantify the local scalings. In the wavelet transform, the contributions of the polynomial trends can be removed effectively. A multi-fractal series can be decomposed into many subsets characterized by different local Hurst exponent h, which quantifies the local singular behavior and thus relates to the local scaling of the series. The statistical properties of these subsets can be quantified by the fractal dimension $D(h)$ of the subset whose local Hurst exponent is h.

As a standard procedure, we first find the WT maximal values, $(T_g(a, \rho_k(a)), k = k_1, k_2, \cdots, k_J)$, where a is the given scale. The partition function should scale in the limit of small scales as,

$$Z(a, q) = \sum_{k=k_1}^{k_J} |T_g(a, \rho_k(a))|^q \sim a^{\tau(q)}. \tag{2.9}$$

The fractal dimension $D(h)$ can be obtained through the Legendre transform,

$$D(h) = qh - \tau(q), \qquad h = \frac{d\tau(q)}{dq}. \tag{2.10}$$

For a mono-fractal structure we have a linear relation, $\tau(q) = qH - 1$. H is the global Hurst exponent. For positive and negative q, $\tau(q)$ reflects the scaling of the large fluctuations and small fluctuations, respectively.

We use the real analytic wavelet $g^{(n)}$ among the class of derivatives of the Gaussian function, by which the polynomial trends up to n order can be removed. The results with $n = 7$ are presented. $n = 5$ and $n = 6$ lead almost the same results. As comparison we detect also the scaling behaviors in the randomized series ρ_R, called shuffled series.

In this thesis, we are interested in the characteristic point at which the fractal dimension reaches its maximum value, $(h_c, D(h_c))$. It can tell us the non-homogeneous distribution of the series ρ and the fractal characteristics of the principal subset.

2.1.5 Numerical Results

We examine the localization behaviors for the cellular networks [90], which are compiled by using a graph-theoretical representation of all the biochemical pathways based upon the WIT integrated-pathway genome database of 43 species from Archaea, Bacteria and Eukarya [91]. The whole cellular networks consider the cellular functions as intermediate metabolism and bioenergetics, information pathways, electron transport, and transmembrane transport. The direct edges are replaced simply with non-directed edges. We consider only the cellular networks with the sizes larger than 500.

We study also the localization behaviors for the the Watts-Strogatz small-world (WSSW) [2] and the Barabasi-Albert scale-free (BASF) [3] networks. For the WSSW model, we construct firstly a regular circle lattice with each node connecting with its d right-handed nearest neighbors. For each edge we rewire it with probability p_r to another randomly selected node. Self- and double-edges are forbidden. By this way, we can introduce randomness into the resulting networks. Moreover, compared with that for the initial regular lattice, the rewiring procedure may introduce also "long-range" edges to the resulting networks, which can reduce significantly the average number of hops required for the nodes to reach each other. This is the so-called small-world effect.

The BASF networks are the results of a preferential growth mechanism which

exists widely in diverse fields. Starting from several connected nodes as a seed, at each growth step a new node is added and w edges are established between this node and the existing network. The probability for an existing node being connected with the new node is proportional to its degree. Self- and double-edges are forbidden. For the resulting networks, the number of edges per node obeys a power-law, namely, no characterized scale exists in this distribution.

Figure 2.1 presents the localization quantities, (Q, S_{str}), for the networks. For the WSSW networks, the randomness introduced by the rewiring procedure has two competitive effects, the long-range edges which favors the extension, and the broken of symmetry which induces the localization. For the BASF networks, the increase of w increases the heterogeneity and the connections between the nodes, which induce the localization and the extension, respectively. Hence, there exist complex relations between p_r or w and (Q, S_{str}) for the two kinds of networks, as shown in Fig. 2.1(a)-(b) and (c)-(d), respectively. For the WSSW networks, the participation ratio decreases rapidly from 1 to 0.22 when p_r changes slightly from 0 to 0.02, and then goes up gradually with the increase of p_r. As for the structural entropy, it increases abruptly when p_r changes from 0 to 0.02, after that it decreases gradually with the increase of p_r.

Figure 2.2 shows S_{str} versus Q. As references, we calculate also the localization quantities for the critical and the localized states, namely, $P(\rho) \sim \rho^{-(1+\eta)}$ and $P(\rho) \sim \frac{1}{\rho}$, respectively. Starting from $\rho(r) \sim r^{-\sigma}$, we calculate the values of $\rho(\frac{n}{N}), n = 1, 2, \cdots, N$. The resulting normalized values can be regarded as the localized state. The critical sates with $\sigma = 1 \sim 10$ are calculated, and the corresponding values of η are $0.5 \sim 0.05$, respectively. N is the size of the considered networks. The same procedure can be used to generate the localized states by

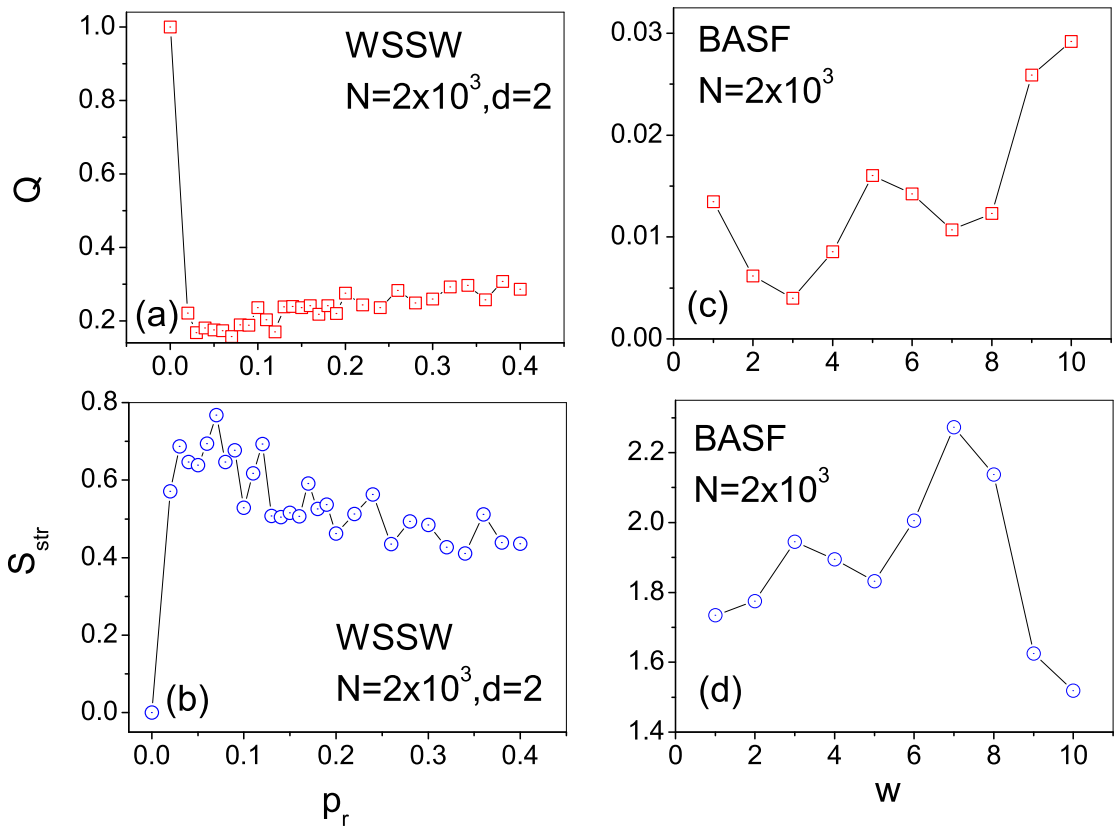

Figure 2.1: **The localization quantities (Q, S_{str}) for the WSSW and BASF networks.** There exist complex relations between p_r or w and (Q, S_{str}) for the BASF and WSSW networks. For the WSSW networks, from $p_r = 0$ to $p_r = 0.02$ there exists an abrupt decrease/increase in value of Q/ S_{str}, as shown in (a)-(b)) respectively. Then with the increase of the rewiring probability p_r the participation ratio tends to increase while the structural entropy tends to decrease; (c)-(d) Results for the BASF networks.

starting from $\rho(r) \sim exp(-\mu r)$. The localized states with $\mu = 0.01 \sim 100$ are generated.

The localization properties of the BASF networks can be captured by the critical states with extremely small values of η. The WSSW and whole cellular networks are generally in between the two typical (localized and extended) states.

We find that the PDFs of the NNLS for all the networks can be described very well by using the Brody distribution in a unified way. The results for the parameter β are shown in Fig. 2.3. For the WSSW networks, with the increase of the rewiring probability p_r, the parameter β increases rapidly from 1.02 ± 0.053 at $p_r = 0$ to 1.95 ± 0.065 at $p_r = 0.14$. For $p_r > 0.14$, β are almost same, namely ~ 2.0. That is, the representative eigenvector changes from a nearly localized state ($p_r = 0$) to an extended state in this interval of p_r. For the networks with $p_r > 0.14$, the representative eigenvectors are almost perfectly extended. While for the BASF networks, with the increase of w, the more edges can induce the significant extensions of the representative states. β reaches its asymptotic value ~ 1.90. Due to the heterogeneity, the BASF networks can not reach a perfectly extended state.

In a considerable wide range of q, the partition functions behave scale-invariant as in Eq. 2.9. There are three kinds of typical WT transform results. Here we present several typical examples. In the whole range of $q = -5 \sim 5$, the WSSW network with $p_r = 0.05$ and the BASF network with $w = 8$ are multi-fractal with only one characteristic point $(h_c, D(h_c))$, as shown in Fig. 2.4 and Fig. 2.5, respectively. Sometimes, the multi-fractal degenerates to mono-fractal. Fig. 2.6 gives another condition where the fractal behaviors can be separated into two branches, namely, $q < 0$ and $q > 0$. The characteristic points for these two

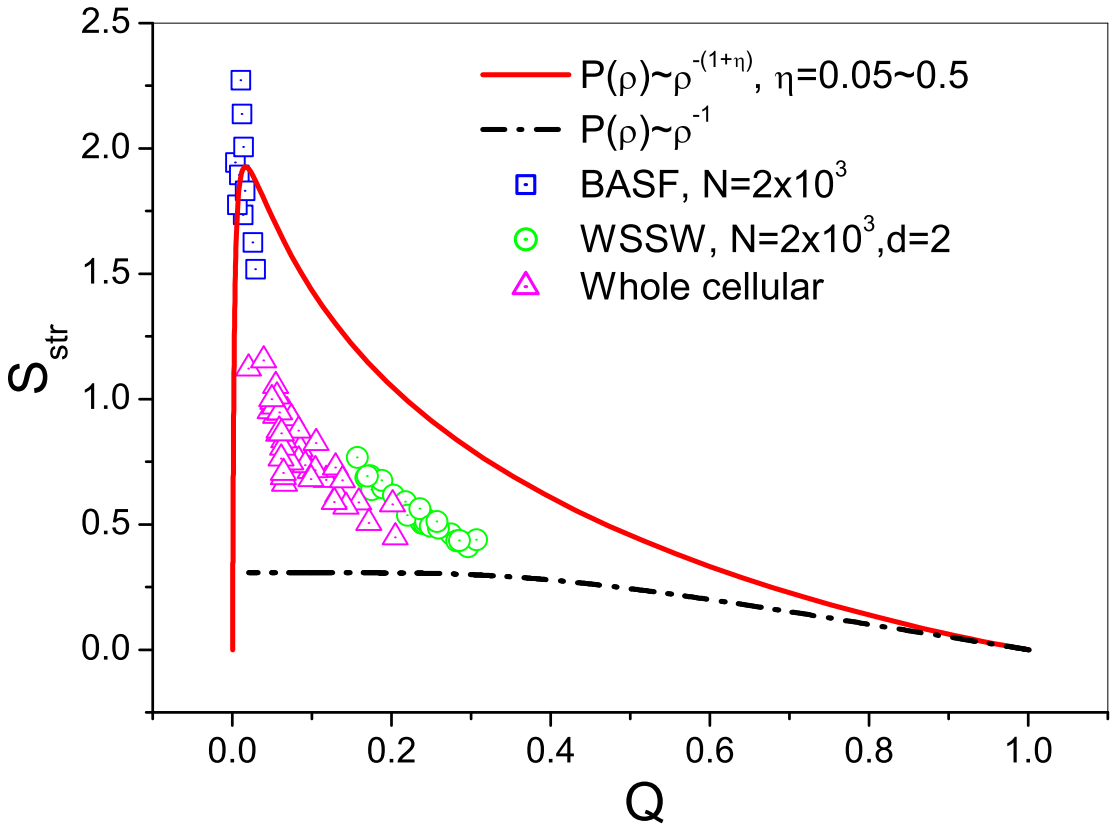

Figure 2.2: **The structure entropy S_{str} versus participation ratio Q for the WSSW, BASF and whole cellular networks.** The localization quantities for the distributions, $P(\rho) \sim \rho^{-(1+\eta)}$ and $P(\rho) \sim \frac{1}{\rho}$, namely, the critical and localized sates, are shown as references. Starting from $\rho(r) \sim r^{-\sigma}$, the set of normalized values of $\rho(\frac{n}{N}), n = 1, 2, \cdots, N$ can be regarded as the critical state. Assigning $\sigma = 1 \sim 10$, the corresponding values of η are $0.5 \sim 0.05$. N is the size of the considered networks. The same procedure is also used to generate the localized states by starting from $\rho(r) \sim exp(-\mu r)$. The localized states with $\mu = 0.01 \sim 100$ are generated. The localization properties of the BASF networks can be captured by the critical state with extremely significantly small values of η. The WSSW and whole cellular networks are generally in between the two typical states.

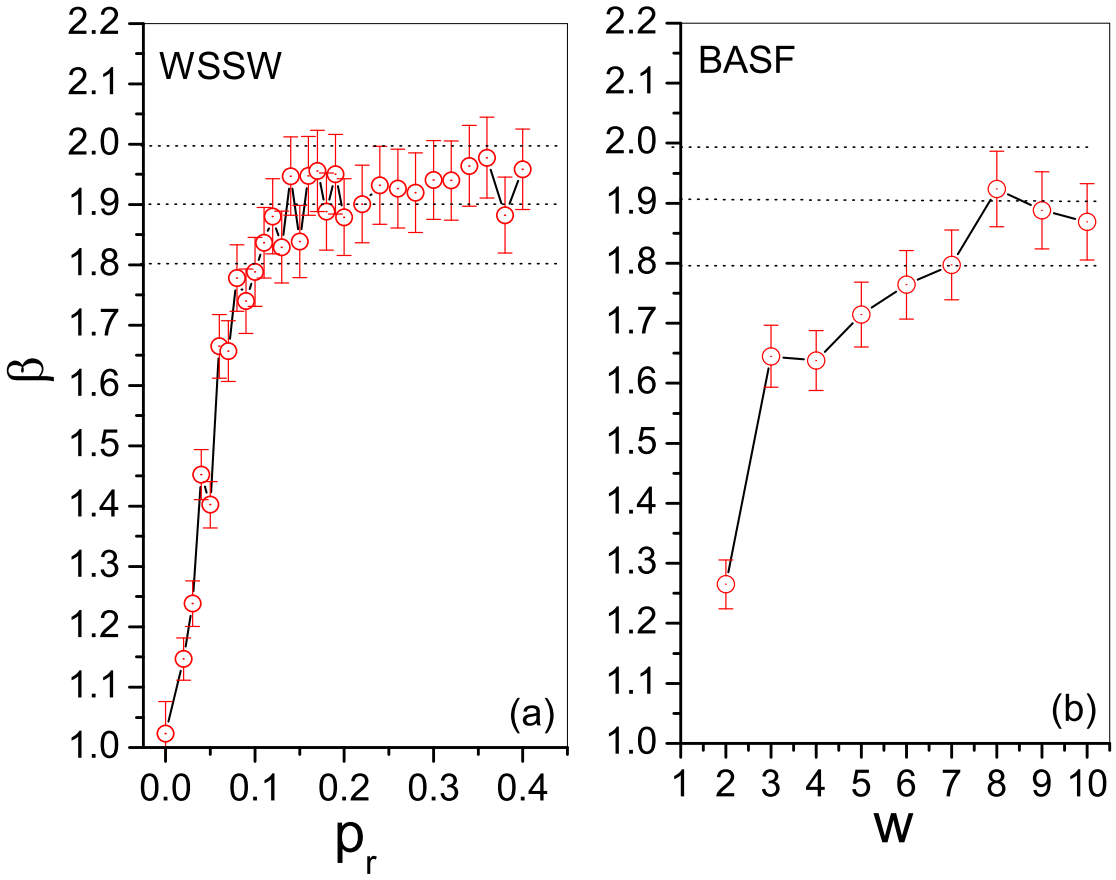

Figure 2.3: **The value of Brody parameter β versus network parameters p_r and w.**(a) WSSW networks; (b) BASF networks.

branches are not same. That is, the principal subsets for the large fluctuations and the small fluctuations are different. These three conditions are called mono-fractal, multi-fractal and branched multi-fractal, respectively.

The scaling properties for the real world networks and the modeling networks are listed in Table 2.1. For the mono- and multi-fractals, we present simply the global Hurst exponent H and the characteristic point $(h_c, D(h_c))$, respectively. For the branched multi-fractal we give the scaling characteristics for the two branches $q < 0$ and $q > 0$, which are separated by the division symbol "/". To cite an example, for the cellular network *M.jannaschii*, the characteristic point for the branch $q < 0$ is $(0.63, 0.96)$ and that for the branch $q > 0$ is $(0.83, 1.03)$. It is denoted with $(0.63, 0.96)/(0.83, 1.03)$. The results for the corresponding shuffled series are presented also. We discard the networks that the scaling behaviors of the original ρ and the randomized series ρ_R are undistinguishable. The sizes of the WSSW and BASF networks are $N = 2000$. $N = 1000, 3000$ and 4000 lead almost same results (not shown in Table.2.1).

The WSSW and BASF networks are almost all mono- or multi-fractals with the values of h_c mainly in the range of 0.66 ± 0.05. However, most of the considered real world networks behave branched multi-fractal. The Hurst exponents lager than 1 and near 0 correspond to non-singularity and white noises, respectively. Discarding these trivial conditions, we find that the multi-fractal behaviors are embedded in the branches of $q > 0$. And the values of h_c are basically in the range of 0.8 ± 0.05. The larger values of h_c for the large fluctuations in the real world networks show us the much more non-homogeneous structures of the PDF of ρ. That is, the PDF of ρ for the real world networks tend to form much sharper peaks at different scales.

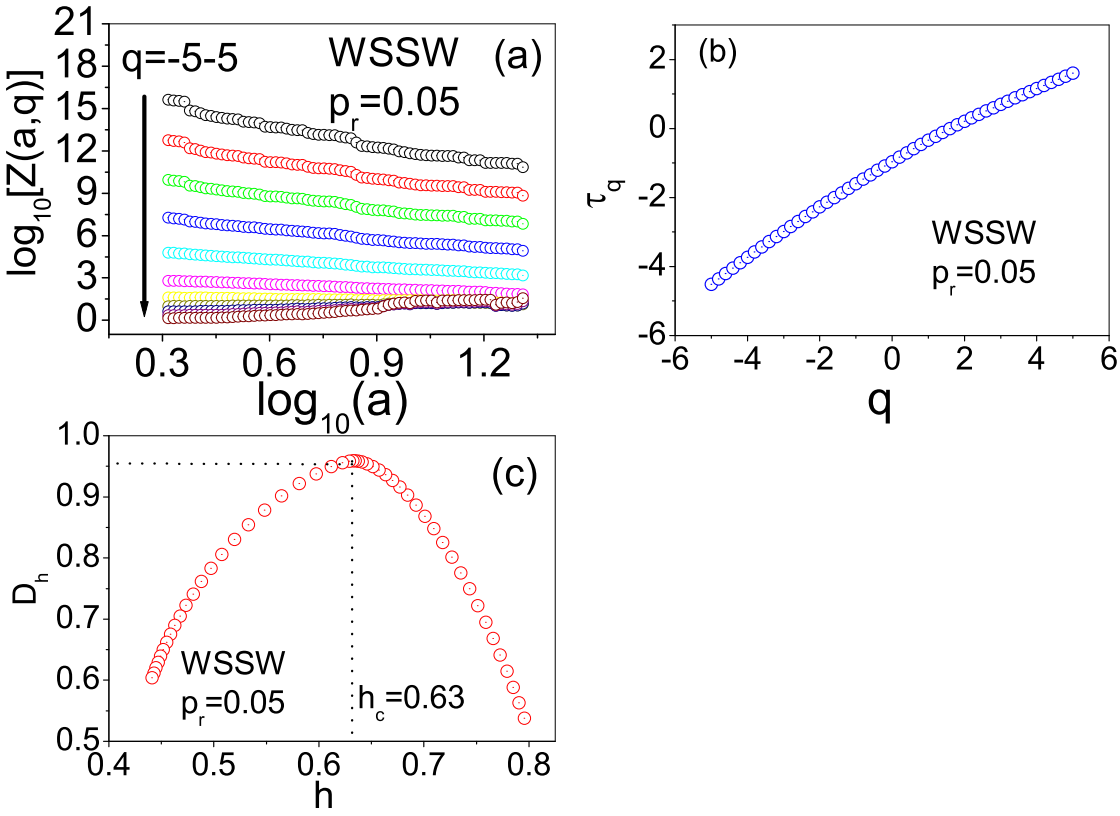

Figure 2.4: **The multi-fractal scaling characteristics of the ascend-ranked series ρ for the WSSW networks.** The multi-fractal behavior for the WSSW network with $d = 2, N = 2000$ and $p_r = 0.05$ is presented as a typical example. In the whole range of $q = -5 \sim 5$, there is only one characteristic point, $(h_c, D(h_c)) = (0.63, 0.958)$.

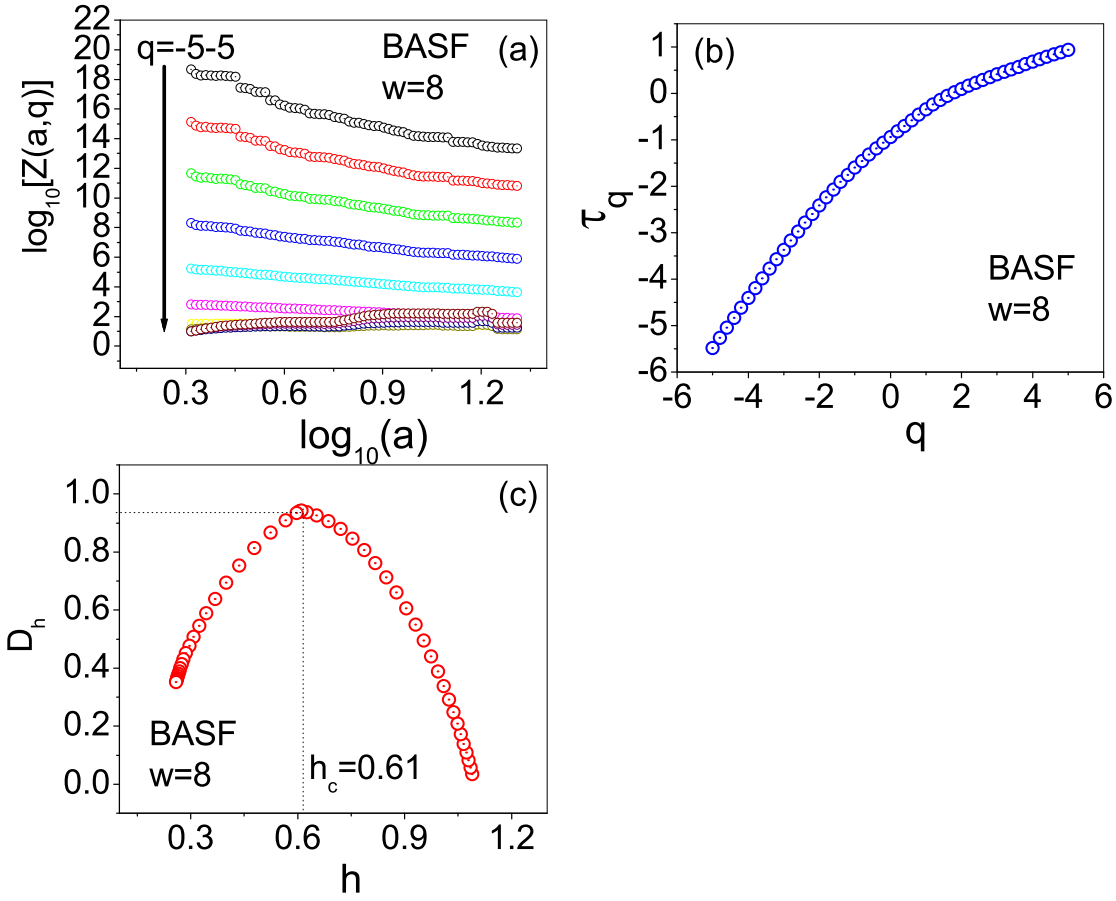

Figure 2.5: **The multi-fractal scaling characteristics of the ascend-ranked series ρ for the BASF networks.** The multi-fractal behavior for the BASF network with $w = 8$ and $N = 2000$ is presented as a typical example. In the whole range of $q = -5 \sim 5$, there is only one characteristic point, $(h_c, D(h_c)) = (0.61, 0.942)$.

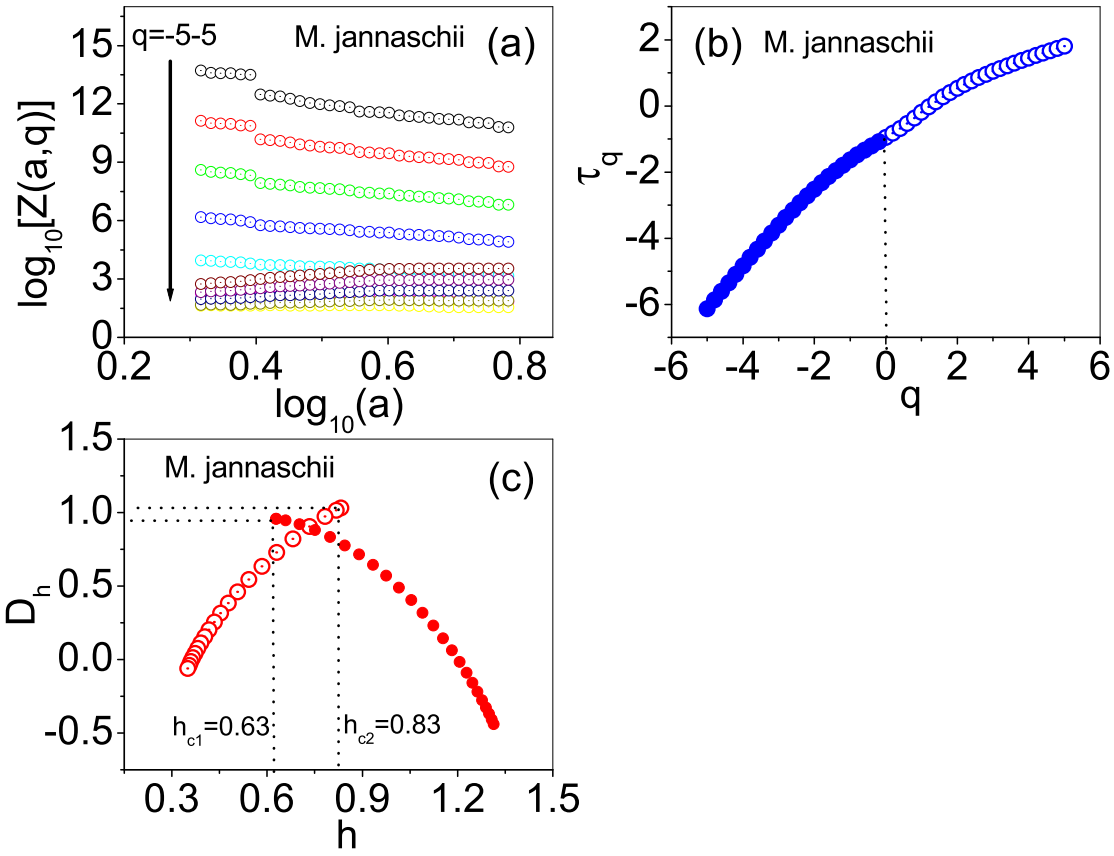

Figure 2.6: **The branched multi-fractal scaling characteristics of the ascend-ranked series ρ for the real world networks.** The branched multi-fractal behavior for the whole cellular network of *M.jannaschii* is presented as a typical example. The two branches $q < 0$ and $q > 0$ lead to different characteristic points, $(h_{c1}, D(h_{c1})) = (0.63, 0.96)$ and $(h_{c2}, D(h_{c2})) = (0.83, 1.03)$.

It should be noted that in the present thesis the structure-induced localization is used as a probe of structure properties of complex networks. We detect the localization properties for the WSSW, BASF model networks and the cellular networks, but it dose not imply and require any localization-related dynamical process (such as waves) occurring on the real-world systems.

Table 2.1: **The scaling properties of the ascend-ranked series ρ for the WSSW, BASF and whole cellular networks.** The characteristic points $(h_c, D(h_c))$ for multi-fractal or H for mono-fractal are presented, which reflect the characteristics of the principal subsets. For the branched multi-fractals, the characteristic points for the branches $q > 0$ and $q < 0$ are separated by the division symbol "/". The results for the shuffled series are also presented as comparison. The h_c for the WSSW and BASF networks are basically in the range of 0.66 ± 0.05, while that or the whole cellular networks are much larger as 0.8 ± 0.05.The asterisks denote the absence of the corresponding branches.

	Networks	Orig series $H, H_1, H_2,$ $(h_c, D(h_c))$	Shuf series $H, H_1, H_2,$ $(h_c, D(h_c))$		Networks	Orig series $H, H_1, H_2,$ $(h_c, D(h_c))$	Shuf series $H, H_1, H_2,$ $(h_c, D(h_c))$
Cellular[3]	*M. jannaschii*	$(0.63, 0.96)/$ $(0.83, 1.03)$	$*$ $*$ $/(0.42, 0.869)$	Cellular[3]	*S. pneumoniae*	0.40	$*$ $*$ $/(0.60, 0.908)$
	A. aeolicus	$0.00/(0.61, 0.94)$	$(0.45, 0.877)$		*M. thermoautotrophicum*	0.72	$(0.48, 0.863)$
	B. subtilis	$3.25/(0.81, 0.95)$	$(0.72, 0.924)$		*T. maritima*	0.33	$*$ $*$ $/(0.66, 0.879)$
	C. acetobutylicum	$2.40/(0.88, 0.99)$	$*$ $*$ $/(0.48, 0.922)$	WSSW	$p_r = 0.02$	$(0.64, 0.980)$	$(0.35, 0.964)$
	C. jejuni	$0.07/(0.82, 0.95)$	$(0.71, 0.883)$		$p_r = 0.06$	$(0.64, 0.959)$	$(0.43, 0.966)$

	E. coli	1.45/(0.78, 0.95)	(0.45, 0.974)/ (0.68, 0.934)		$p_r = 0.10$	(0.62, 1.004)	(0.34, 0.963)
	M. bovis	2.82/(0.85, 0.88)	(0.42, 0.889)		$p_r = 0.14$	(0.68, 0.956)	(0.33, 0.963)
	M. tuberculosis	4.04/(0.75, 0.98)	(0.68, 0.936)		$p_r = 0.18$	(0.64, 0.989)	(0.34, 0.964)
	N. meningitidis	0.22/(1.20, 1.06)	(0.76, 0.907)		$p_r = 0.24$	1.000	(0.32, 0.967)
	S. cerevisiae	2.54/(0.89, 0.92)	$*$ $*$ /(0.61, 0.889)		$p_r = 0.28$	0.714	(0.32, 0.964)
	A. fulgidus	0.95/1.23	$*$ $*$ /(0.57, 0.903)		$p_r = 0.32$	(0.61, 0.967)	(0.27, 0.963)
	C. elegans	5.70/0.71	$*$ $*$ /(0.50, 0.897)		$p_r = 0.36$	0.687	(0.33, 0.962)
	C. tepidum	1.97/0.20	(0.49, 0.928)		$p_r = 0.40$	0.546	(0.28, 0.961)
	D. radiodurans	2.82/1.14	(0.75, 0.964)	BASF	$w = 1$	1.099	0.887/0.00
	H. influenzae	3.82/0.59	(0.81, 0.964)		$w = 2$	(0.58, 0.976)	1.500/0.00
	H. pylori	4.33/0.81	(0.61, 0.895)		$w = 3$	0.638	1.135/0.00
	M. leprae	3.68/0.50	(0.70, 0.887)		$w = 4$	1.06/0.40	1.197/0.00

	N. gonorrhoeae	2.84/1.13	(0.60, 0.923)		$w = 5$	1.15/0.63	0.901/0.00
	P. gingivalis	3.44/0.63	$*$ $*$ /(0.41, 0.912)		$w = 6$	(0.64, 0.961)	0.847/0.00
	R. capsulatus	1.59/0.00	$*$ $*$ /(0.68, 0.902)		$w = 7$	(0.72, 1.014)	0.838/0.00
	P. aerugi-nosa	0.14/0.57	(0.68, 0.915)		$w = 8$	(0.61, 0.942)	0.828/0.00
	A. antino-mycetem-comitans	(0.54, 0.924)	(0.45, 0.901)		$w = 9$	0.516	0.691/0.00
	E. faecalis	(0.84, 0.954)	(0.65, 0.880)		$w = 10$	0.721	0.696/0.00

2.2 Localizations on Directed Networks

As we presented in the first part, all the above investigations are performed for un-directed networks which has symmetric adjacency matrix and whose eigenvalues are real. But many complex systems are described by directed networks [1], for example world wide web, social networks, and PPI biological networks, henceforth lead to complex eigenvalues of the corresponding adjacency matrices.

As we know, most of the theoretical investigations and applications of RMT

have been focused on symmetric matrices. Asymmetric matrices are less well understood. Recently few numerical studies on spectral properties of asymmetric matrices have been done in the context of google networks [92, 93].

In this part, we will give the spectra analysis of directed networks where entries in corresponding matrix take values motivated by inhibitory and excitatory coupling between nodes, and the entries in the matrix take values 0, -1 and $+1$, respectively. The matrix corresponding to these networks is not *random*, position of entries depends on the underlying network structure. The real world networks and model networks are sparse, so most of the entries in the corresponding matrix are zero.

Spectral densities of directed random networks with entries 0, $+1$ and -1 have been studied analytically and numerically in [94]. In [94], eigenvalues spectra of large random matrices with excitatory and inhibitory columns drawn from distribution with different means and equal or different variances have been computed. Interestingly, spectra of these random matrices (with 0 and 1, -1 entries) also follow Girko's circle law.

In this section, we will investigated spectra and localization properties of directed networks with binary entries.

2.2.1 Spectra Analysis Methods

We obtain directionality of edges of networks as following: assign equal probability to a node to be inhibitory or excitatory. All edges starting from an excitatory node would take value $+1$, and those starting from an inhibitory node would take value -1. It means the corresponding adjacency matrix A gets entries as follows: if ith node is excitatory, $A_{ij} = 1$ for all the j nodes which have connections with

i, and if ith node is inhibitory, $A_{ij} = -1$. The corresponding matrix has following two properties: (a) $|A_{ij}| = |A_{ji}|$, and (b) a row has all entries either $+1$ or -1. This configuration would lead to zero mean ($\sum_{i,j} A(i,j) \sim 0$), and standard deviation $\sigma^2 \sim \tau$. We denote the eigenvalues of network by $\lambda_n,\ n = 1, \ldots, N$, where N is the size of networks.

Then we numerically analyze the adjacency matrix of the network to obtain a set of eigenvalues $\lambda_m = R_m + iI_m$, and eigenvectors V_i. To characterize localization properties of eigenvectors V_i, we use Inverse Participation Ratio (IPR) which is the inverse of the Participation Ratio Q we used in Section 2.1.2. It is defined as,

$$I_i = \frac{(\sum_j |V_i(j)|^2)^2}{\sum_j |V_i(j)|^4}, \tag{2.11}$$

where $V_i(j), j = 1, \ldots, N$ are the components of eigenvector V^i. The meaning of I is illustrated by two limiting cases : (i) a vector with identical components $V_i(j) \equiv 1/\sqrt{N}$ has $I_i = N$, means most extended, whereas (ii) a vector with one component $V_i(j) = 1$ and the remainders zero has $I_i = 1$, here the eigenstate is localized.

Although the IPR in general is used to understand the localization property of individual eigenstates, the sum of all I_i ,

$$< I >=< \sum_i I_i >, \tag{2.12}$$

where angular bracket denotes an ensemble average, can be used as a measure for localization properties of whole spectra altogether [95].

2.2.2 Spectral Properties for Completely Uncorrelated (Directed) Random Networks

In the following, first we provide the eigenstate properties of Erdös-Rényi random networks [96] having different average degrees, and then track the spectral properties as a parameter of τ. The value of τ is a measure of uncorrelaion (directionality) or correlation for a network.

We consider completely uncorrelated random networks, for which $\tau = 0$ and connection probability is p. Since we take equal probability of a node being inhibitory or excitatory, the expected number of $+1$ and -1 entries are approximately same. The mean and variance of this network can be calculated as $\mu = 0$ and $\sigma^2 = p$.

Fig. 2.7 plots eigenvalues together with IPR for random networks with different probabilities of connections.

As we do not consider self-connection, the diagonal elements of the matrix are zero, and hence eigenvalues are distributed around origin 0. Eigenvalues are distributed in circular region of radius $\sqrt{Np(1-p)}$. For larger values of p, few pairs of eigenvalues get separated from the circular bulk. The bulk circular region still lie between $-\sqrt{Np(1-p)}, \sqrt{Np(1-p)}$. For $p = 1$ which corresponds to globally coupled network, there is one conjugate pair of eigenvalues with absolute value $\sqrt{N}$ which are non-degenerate, rest of eigenvalues are degenerate with values $+1$ and -1.

Different connections probabilities lead to different expected average degree of the network. For low connection probability, $p = 0.005$, one gets localized eigenstates at four corners of the real and imaginary axis. Apart this four corners

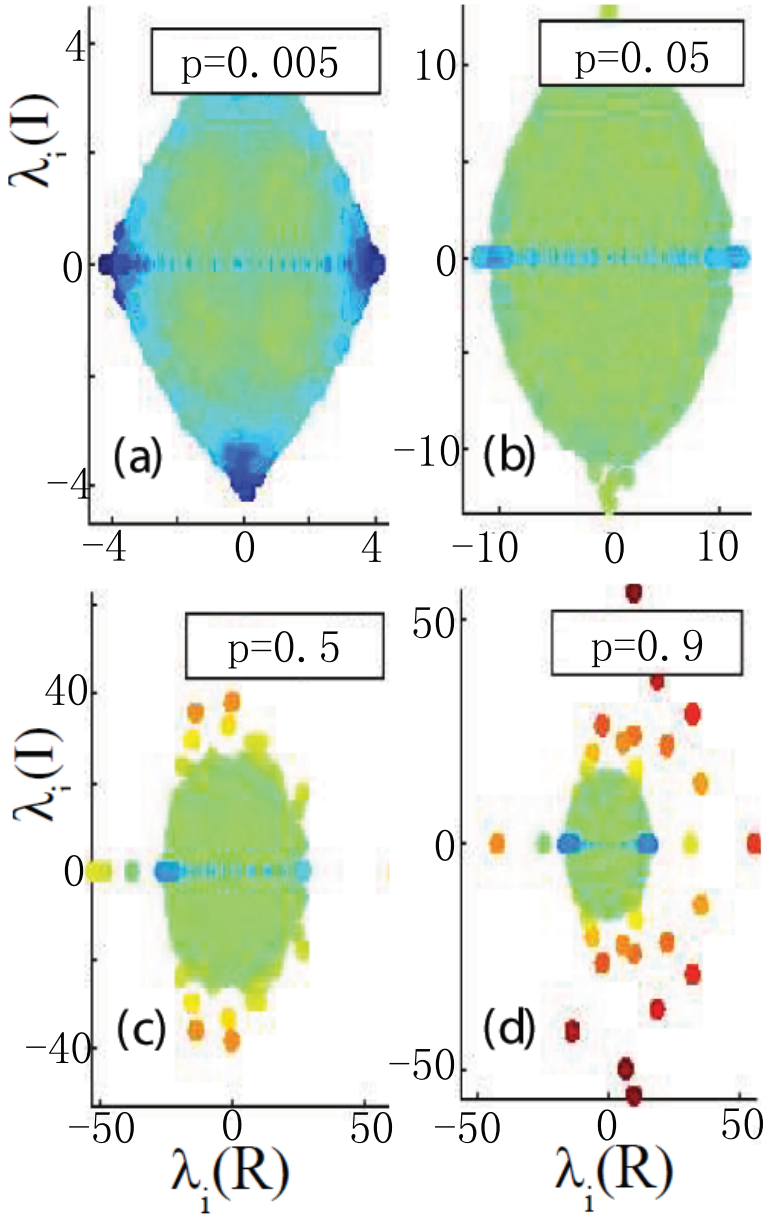

Figure 2.7: **Spectra with IPR for directed random networks having different connection probabilities** p**.** Size of networks are $N = 2000$, and spectra are plotted for 10 random realization for each value of p. x and y axis plot the real and imaginary part of the eigenvalues, and color denotes the value of corresponding IPR. Value of IPR varies from $IPR = 1$ (localized, dark gray blue) to $IPR = 2000$ (extended, dark gray red). Note the scale of color axis for this case, which is different from other figures.

the eigenstates on real axis are also more localized. As connection probability increases overall distribution of eigenvalues remains same, i.e. homogeneous, except for larger p few pairs of eigenvalues get separated from the bulk circular region [92]. Except eigenstates lying on real axis, other eigenstates located towards light gray blue, denoting delocalization. The isolated eigenstates are maximal delocalized, which implies that eigenstates having large absolute eigenvalues are more random than the bulk part. Very small values of p yields a delocalized spectra (Fig 2.7a), which may be due to the sparseness of connections.

Here we would like to mention that, the above behavior is for the special arrangement of $+1$ and -1 entries in the matrix, a row has all $+1$ or -1 entries depending upon whether the corresponding node is excitatory or inhibitory. For a random network with entries $+1$ and -1 randomly distributed, the radius of circular region scales with average degree of the network, i.e. $\sqrt{pN}$, and all eigenvalues lie within the bulk regions even for larger value of p including $p = 1$.

2.2.3 Tracking Spectral Localization Properties from Symmetric to Asymmetric (or Directed) Networks

In this section we investigate the origin of localization or delocalization of eigenstates by tracking the spectra as connections are made undirected. We start with a network having all nodes excitatory leading to a symmetric network, now with probability (1- τ)/2 some nodes are made inhibitory. $\tau = 0$ corresponds to the case when half of nodes are made inhibitory, leading to a completely uncorrelated network. As the fraction τ of connections are made symmetric, the mean and variance take forms as:

$$\begin{aligned} \mu &= \sum_{ij} A_{ij} \sim \tau p, \\ \sigma^2 &= \sum_{ij} (A_{ij} - \mu)^2 \sim p(1 + \tau p - 2\tau^2 p), \end{aligned} \tag{2.13}$$

where $\tau = 1$ corresponds to a symmetric network which we call a complete deviation from the directed network, i.e. $A(i,j) = A(j,i)$. We call "deviation from directionality" because as one deviates from completely directed networks, spectra deviate from the circular structure. For a network, where nodes are inhibitory or excitatory with equal probability, $< A_{ij} >= 0$ and $< A_{ij}^2 > \sim p$.

Figure 2.8 plots the spectra along with IPR values for random networks for different values of τ. The largest eigenvalue lies on real axis well separated from the bulk. Isolation of largest eigenvalue from rest is well known observation for real symmetric random matrices having non zero mean [97]. It was proved in [97] that such an isolated eigenvalue exists for a large matrix if the mean value of the elements is not zero and rather carries substantial fraction of the root mean square of the elements. Further more, the associated eigenvector would have all its components close to equality.

Now let us first consider the case with $\tau = 1$. The corresponding matrix would be a symmetric matrix with $pN(N-1)$ entries being $+1$ and rest being 0, the mean and variance of this matrix can be calculated as $\mu = p$ and $\sigma^2 = p(1-p)$, respectively (Eq. 2.13). The largest eigenvalue of the matrix scales as pN, which is equal to the average degree of the network. Rest of the eigenvalues are homogeneously distributed in a circular region of radius $pN(1-p)$.

As connected are made directed, the mean μ decreases, and for $\tau = 0$, mean takes value 0, and variance is simply given by p which is connection probability. As

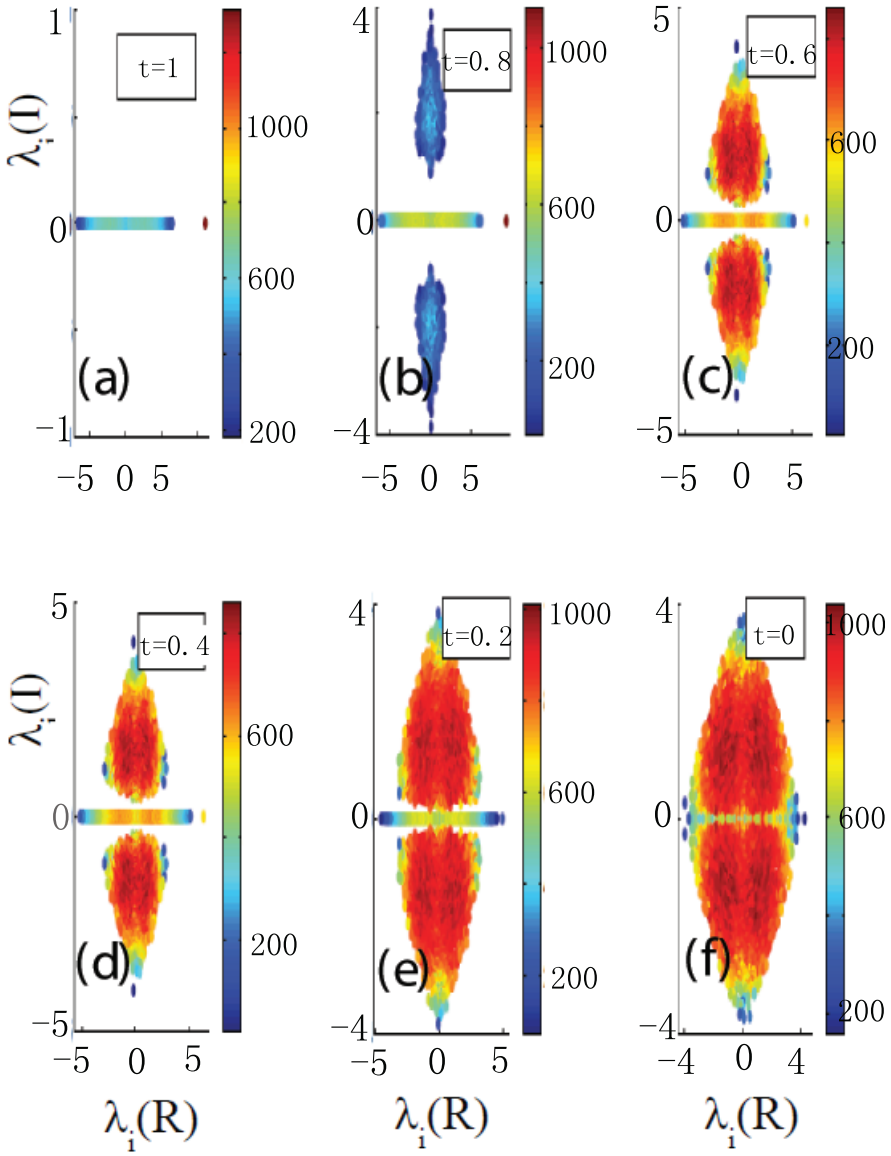

Figure 2.8: **Spectra with the IPR for random networks having different values of τ.** The size and average degree of the networks in all sub-figures are the same which are $N = 2000$ and $< k >= 10$ respectively. $\tau = 1$ corresponds to a symmetric network, and $\tau = 0$ corresponds to complete asymmetric network (see the text). the values of IPR varies from $I = 0$ [dark gray (blue)at the bottom of the shading scale] to $I = 1300$ [dark gray(red) at the top of the scale].

connections are made directed, eigenvalues start appearing in complex conjugate pairs, and we can divide spectra into two parts: (A) part of spectra with real eigenvalues, and (B) part of spectra having complex conjugate pairs of eigenvalues. For $\tau = 0.8$, which means 20% of connections are made directed, a large number of eigenvalues is still in group (A), i.e., they lie on real axis. Complex conjugate pairs of eigenvalues form a oval shape which is well separated from real axis (see Figure 2.8). The localization properties of eigenstates from both the groups do not change much, as bulk part which is separated from real axis is still delocalized, only middle part of eigenstates lying at real axis are more localized (group (A)) delocalized than a completely symmetric network.

As more connections are made directed, complex conjugate pairs of eigenvalues start appearing with delocalized states. Bulk of the complex conjugates pairs lying at both sides of real axis and forming oval shape (group (B)) are delocalized, only few eigenvalues in this group with largest absolute value remain localized. Eigenvalues forming group (A) also show similar features as group (B), eigenstates with larger eigenvalues at both ends remain localized, whereas rest of the eigenstates are more delocalized than earlier.

Moreover, with directed connections increasing (value of τ is decreased), size of group (B) also keeps increasing, and consequently two parts of eigenstates come closer. For $\tau = 0.2$, group (B) is no longer separated from the eigenstates lying on real axis (group (A)).

2.2.4 The Localization Properties for the Whole Networks

We also calculate total $IPR < I >$ in order to get a measure for localization of the whole network.

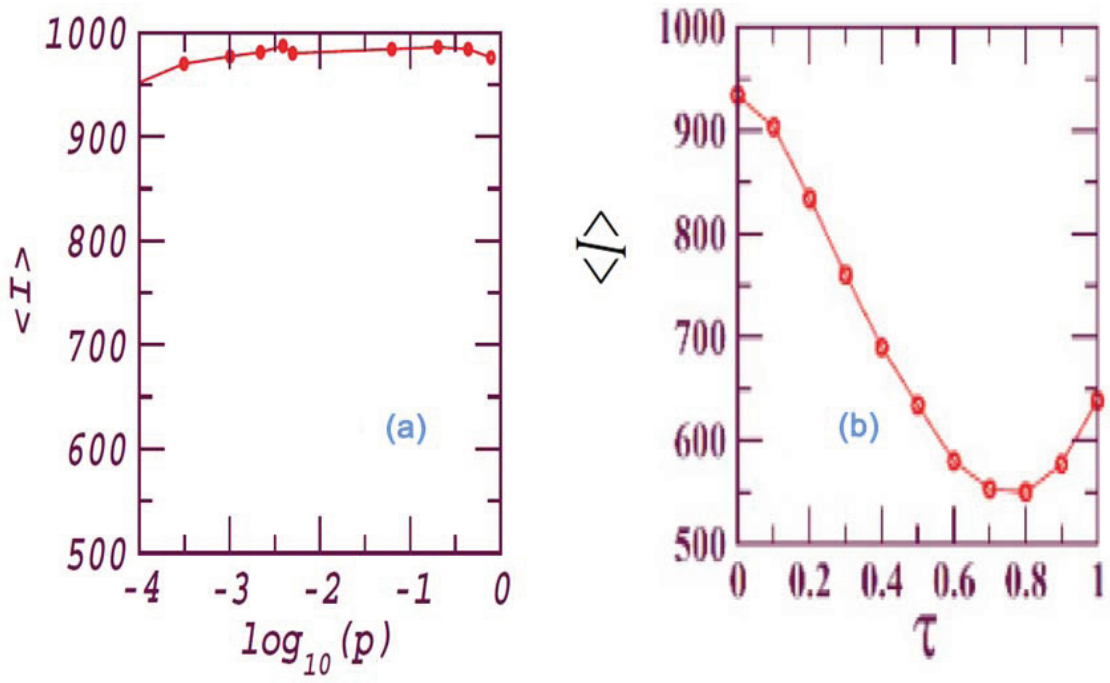

Figure 2.9: **The total IPR for directed random network.** $(a) < I >$ vs $\lg p$ for a random network with $N = 2000$,$\tau = 0$ (completed uncorrelated network) for various value of p or an average degree $< k >$. $(b) < I >$ vs τ for $N = 2000$, $< k >= 20$ and for ten ensemble average of the random networks.

In Fig. 2.9 (a), we plot the total IPR, $< I >$ vs $\lg p$ for completely uncorrelated (directed) random networks mentioned in Sec. 2.2.2, with varying value of p or an average degree $< k >$. The network size always remains same, $N = 2000$. The range for p is taken from $p = 0.001$ to $p = 0.9$. We do not plot the graph below $p \sim 0.001$, as for $N = 2000$, smaller value of p yields several disconnected clusters.

Fig. 2.9 (b) plots the total IPR, $< I >$ vs τ for the networks mentioned in Sec. 2.2.3, with a particular value of connection probability $p = 0.01$ and network size $N = 2000$. Starting from a symmetric network ($\tau = 1$), connection are made directed with probability $(1-\tau)$. Overall spectral delocalization increases as connections are made directed leading to occurrence of complex conjugate pairs of eigenvalues in spectra. Delocalization reaches its maximum for $\tau = 0$ value which corresponds to a complete un-correlated or directed network.

2.3 Summary

In this chapter,we have investigated the localization properties on complex networks.

For the undirected networks, the probability distribution function of the representative eigenvector is proposed to describe the localization properties of complex networks. The localization quantities (Q, S_{str}), the PDFs of the NNLS and the wavelet transform have been used to capture the characteristics of the representative state. The nontrivial structures of the networks can induce the localizations of the representative states. At the same time, because of the global symmetries of the networks, the representative state have nontrivial structures rather than the

step-like distribution.

In addition, we have also analyzed spectra and localization properties of directed networks with binary entries. The networks with inhibitory and excitatory nodes have much rich spectra than the undirected networks. The particular eigenvector localization properties of random networks for different values of correlation among their entries were investigated. Spectra of random networks with completely uncorrelated entries show a circular distribution with delocalized eigenvectors, whereas networks with correlated entries have localized eigenvectors. In order to understand the origin of localization, the spectra as a function of connection probability and directionality were tracked here. As connections are made directed, eigenstates start occurring in complex-conjugate pairs and the eigenvalue distribution combined with the localization measure shows a rich pattern.

Chapter 3

Evolutionary Clues Embedded in Network Structure

In the last chapter, we have studied the large scale structure and systematic level dynamics of certain model networks and real world networks using tools from Random Matrix Theory(RMT) and nonlinear dynamics.

In this chapter, we present a spectral-analysis based method to detect the evolutionary history of a network which is a fundamental question in network science. We find that in a complex network, different groups of nodes may have been existed for different amounts of time. In particular, we find that there are complex networks in the real-world for which there is a positive correlation between the eigenvalue magnitude and node age.

First, we concentrate on the situations where the network topology has already obtained. Then for the unknown structure networks, but short time series measured from nodes are available, we suggest to uncover the network topology at the present (or any given time of interest) by using compressive sensing and then perform the

spectral analysis.

This chapter is based on my third publication Ref.[98].

3.1 Motivations

Many large complex networks in existence today are the results of some evolutionary processes such as growth [99]. The Internet is one best example, which has undergone tremendous expansion in the past two decades. Growth in a decentralized manner also appears to be the hallmark of other types of networks such as various biological, social and economical networks (e.g., Facebook). Given a complex network but without any knowledge of its evolutionary history, one might be interested in the distribution of the "ages" of various nodes or subgroups of nodes in the network. Information about the node ages can provide deep insights into the organization and structure of the underlying network, and may have significant applications. For example, in a social network, the lifetimes of certain subgroups of nodes may be closely related to the network backbone structure in terms of the roles that these subgroups play in the function of the network, e.g., leadership roles. In a biological network, nodes of longer lifetimes can be more critical to the various functions of the network. It is thus of considerable interest to develop a systematic method to uncover the evolutionary ages of subgroups of nodes in complex networks.

Two situations arise when addressing the age-detection problem in complex networks: (1) network topology is known and (2) the topology is unknown but only time series measured or observed from various nodes are available.

In the first case we shall establish that the spectrum of the network connectivity matrix, or the Laplacian matrix, is directly related to the evolutionary ages of various subgroups of nodes in the network. In the second case, we shall make use of a recently developed method of time-series based reverse engineering of complex networks [100] to uncover the network topology, and then could analyze the spectrum of the predicted Laplacian matrix to obtain estimates of the age distribution of nodes. Our approach thus defines a framework in which the problem of evolutionary-age detection of nodes in complex networks can be addressed in a systematic way. While our method does not require a positive correlation between the node degree and age, a correlation between the eigenvalue and the node age is necessary.

It is useful to point out that for the class of scale-free networks that are generated according to the preferential-attachment rule [3], the problem of evolutionary-age estimation may be trivial. In particular, this growth rule stipulates that the probability for an existing node to acquire new links is proportional to its degree, implying a strong correlation between the node degree and its lifetime. Thus, for a scale-free network evolved predominantly according to the preferential-attachment rule, the ages of various nodes can be predicted simply by examining the degrees. However, many real-world networks deviate significantly from the scale-free topology [99] and, for them the problem of detecting node evolutionary ages is nontrivial. Nonetheless, scale-free networks provide an ideal test bed to validate our spectrum-analysis method.

We shall emphasize that, although our method is suitable even for networks for which there is no positive correlation between node degree and age, its applicability is limited to networks for which there is a positive correlation between the

properties of the eigenmodes and the node age. For networks with which no evolutionary process can be affiliated, such as various citation networks and twitter-type of social networks where the importance of a node may not be related with its age, our method is not applicable.

In Sec. 3.2, we shall describe the main ideas underlying our method. In Sec. 3.3, we shall validate the method by using scale-free networks generated by the standard preferential-attachment rule and by the duplication/divergence mechanism, which are especially relevant to social and biological systems, respectively. In Sec. 3.4, we shall consider a realistic biological network, the protein-protein interaction network for which the age distribution of nodes is available, to further validate our method. In Sec. 3.5, we shall address the situation where the network topology is not known *a priori* but only time series are available. We shall make use of the reverse-engineering approach [100] to map out the network topology, and demonstrate that our method yields correctly and accurately the spectrum of the Laplacian matrix. A brief conclusion is presented in Sec. 3.6.

3.2 Method

For an undirected complex network with N identical nodes, its topological structure can be described by the Laplacian matrix L [56, 57, 101–103], where the off-diagonal elements of L are $L_{i\neq j} = L_{j\neq i} = -1(0)$, if the nodes i and j are connected (disconnected), respectively. The diagonal elements are $L_{ii} = -\sum_{j\neq i} L_{ij} = k_i$, where k_i is the number of the nodes connected directly with the node i (node degree).

We consider the nodes as atoms and the edges as bonds, the network can

be mapped to a large molecule. For an electron moving in such a molecule, the tight-binding Hamiltonian of the system reads,

$$\mathcal{H} = \sum_{n=1}^{N} \varepsilon_n \cdot |n\rangle \langle n| + \sum_{m \neq n}^{N} \tau_{mn} \cdot |m\rangle \langle n|, \tag{3.1}$$

where ε_n is the site energy and τ_{mn} is the hopping integral for the bond between sites m and n .

Because all the nodes and the couplings are identical, we have

$$\tau_{m \neq n} = cL_{m \neq n}. \tag{3.2}$$

The parameter c is the coupling strength for each bond. Without lose of generalization, we can set $c = 1$. We assign the site energy, $\varepsilon_i = k_i$. These two assumptions lead to

$$H = L. \tag{3.3}$$

For a connected network, the eigenvalues of L are nonnegative and can be ranked as $0 = \lambda_1 \leq \lambda_2 \cdots \leq \lambda_N$. The corresponding eigenvectors are $X_1, X_2, \cdots, X_N$, whose wavelengths are sorted in a descending order. Each eigenvector contains components concentrated on various nodes in the network.

For a regular or a small-world network [2], the eigenvectors typically exhibit some wave patterns with certain wavelengths [104, 105]. When a perturbation is applied to the network, the affected eigenvectors are those whose wavelengths match the size of the perturbation (i.e., the number of nodes that it affects). In this case, some localized structure in the affected eigenvectors can emerge. Eigenvectors associated with small eigenvalues usually have large wavelengths, and so they are sensitive to perturbation on a global scale. In contrast, eigenvectors associated with

large eigenvalues are most sensitive to localized perturbations that are applied to a small set of nodes in the network.

An example is given in Fig. 3.1, for a one-dimensional regular lattice of $N = 10$ nodes with periodic boundary condition, where each node is connected with 2 neighbors on either side so that the node has 4 nearest neighbors. Then we choose node 6 connecting to node 3 and node 9. The new network is considered as a perturbed one.

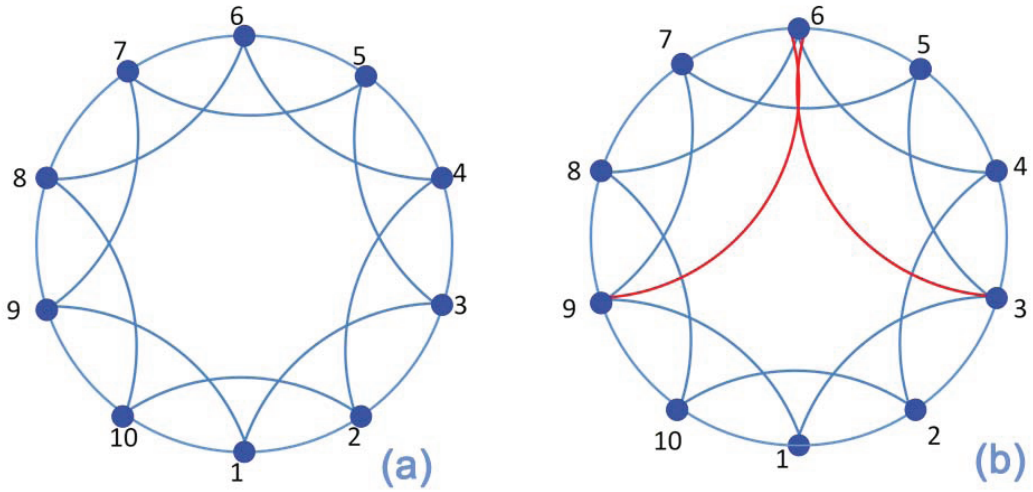

Figure 3.1: **Examples of** (a) Regular ring network $N = 10$, (b) Perturbed network by connecting node 6 with node 3 and node 9.

The responses of the eigenvectors to perturbations thus reflect the structure of the network at different scales. To deep study this point, we give a larger network in Fig. 3.2, which is also for a one-dimensional regular lattice of $N = 100$ nodes with periodic boundary condition, where each node is also connected with 2 neighbors on either side.

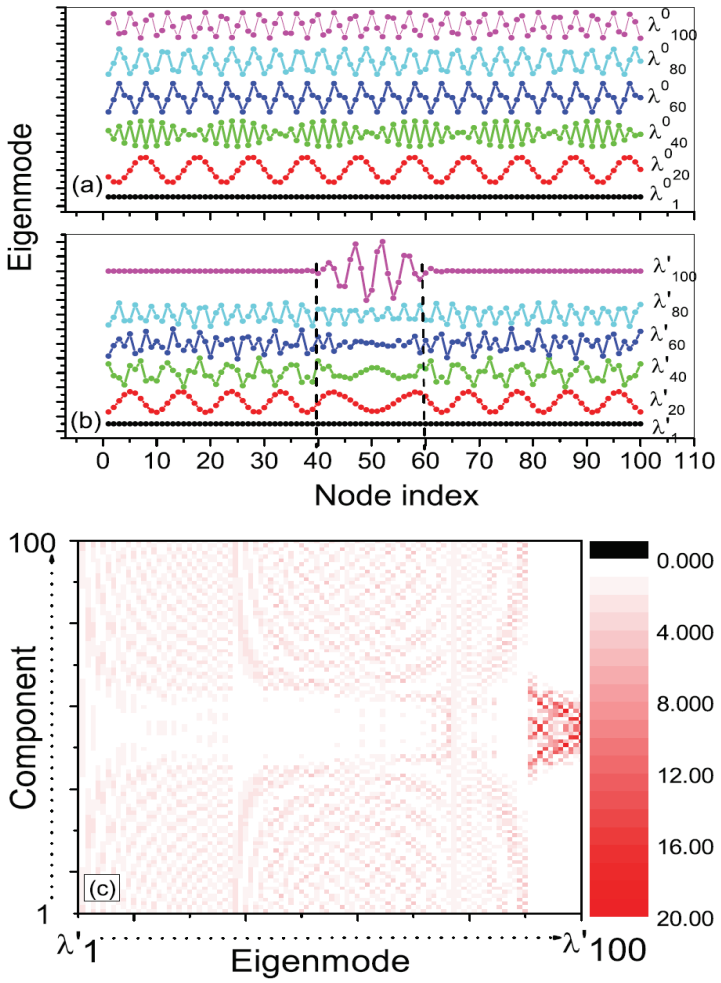

Figure 3.2: **The sensitive localization effects on eigenvectors which are affected by the structure perturbation to the small world networks.** For a regular ring network of 100 nodes where each node has four neighbors, (a) examples of typical, periodic-wave like eigenvectors, (b) typical eigenvectors when each node in the group of indices between 40 and 60 acquires two additional links, one on each side. We observe significant distortions from the periodic-wave pattern, which are localized between the 40th and 60th components of eigenvectors associated with relatively large eigenvalues. (c) Representation of all eigenvectors, where those associated with eigenvalues from λ'_{90} to λ'_{100} are significantly more sensitive to the structural perturbation to the network.

Shown in Fig. 3.2(a) are representative eigenvectors, where the values of $N \cdot X_i^2(s)$ are plotted and $X_i^2(s)$ is the sth component of the eigenvector X_i. We see that the eigenvectors represent periodic waves of wavelengths ranging from N to 2. To observe the effect of local structural perturbation on the eigenvectors, we add two more links to each node in the group of nodes whose indices are between 40 and 60 so that each node in this perturbed group now has six nearest neighbors.

Let λ_i' $(i = 1, \ldots, N)$ be the eigenvalues in the perturbed network. Figure 3.2(b) shows some representative eigenvectors. We observe that the eigenvectors associated with small eigenvalues, e.g., λ_1', λ_{20}', λ_{40}', λ_{60}', and λ_{80}', are basically unchanged. However, eigenvectors associated with relatively large eigenvalues, such as λ_{100}', are strongly altered by the perturbation but the changes are focused on the perturbed group of nodes.

Figure 3.2(c) shows the distribution of the magnitudes of all eigenvectors on nodes in the network, where we see that those associated with eigenvalues λ_{90}' to λ_{100}' are sensitive to the perturbation with large variations appearing on the perturbed nodes.

It should be noted that the eigenmodes are just a simple extension of the normal modes in Gaussian network model (GNM) [106, 107]. In the elastic networks described by GNM, the position of each node can be described with its coordinates in Euclidean space, while complex networks have only topological structures. Because our focus is on the nontrivial properties of eigenmodes induced by structural patterns, we reduce the harmonic couplings to a simple version instead of the 3-dimensional one in GNM. What is more, the eigenmodes are used just as probes of structural patterns. It does not imply and require any relevant dynamical processes occurring in the real-world networks.

For complex networks that do not possess a regular backbone, such as random [96] and scale-free [3] networks, the eigenvectors in general do not exhibit any periodic wave structure. Nonetheless, the observation that the eigenvectors associated with larger eigenvalues are more sensitive to structural perturbations can be used to infer the evolutionary age of nodes. To see this, consider a scale-free network evolved according to the preferential-attachment rule [3], for which there is a positive correlation between the node degree and lifetime. That is, nodes of "old" ages tend to have more links and they are thus more susceptible to perturbations applied randomly to the network during the evolutionary process. Since the eigenvectors of large eigenvalues are quite sensitive to perturbations (c.f., Fig. 3.2), we expect the large-degree nodes to dominate these eigenvectors. As a result, large eigenvalues tend to correspond to nodes of long lifetime. This argument suggests that, nodes having the most significant components of the eigenvectors associated with the largest eigenvalues are likely to possess the longest lifetime in the network.

3.3 Validation with Scale-Free Networks

To exemplify the relation between eigenvalues and node ages, we consider standard scale-free networks [3]. Each network has $N = 2000$ nodes, which is evolved following the preferential-attachment rule so that the age of the ith node is $N - i + 1$. For a given eigenvalue, the lifetime of the associated eigenvector is the average age of all nodes contained in the vector, weighted by the respective components of the eigenvector.

Figures 3.3(a-c) show the ages of the eigenvectors X_i versus the index i for three networks of different edge density w. The significant feature common to

all three cases is that the average age of the nodes dominating some eigenvector increases on average with the eigenvalue. The average degree of each eigenvector, i.e., the weighted average of the degrees of all nodes associated with the vector, shows the same tendency, as shown in Figs. 3.3(d-f), where the average degree is presented on a logarithmic scale.

For each network, the sizes of the eigenvectors are shown in Figs. 3.3(g-i), where the size of an eigenvector is defined to be the number of nodes on which the vector component is larger than a small threshold value. For sufficiently dense network, e.g., Fig. 3.3(i), the size tends to decrease on average with the eigenvalue, indicating that a small group of nodes have extraordinarily long lifetimes in the network and their relative ages can be identified simply by examining the associated eigenvalues. Figures 3.3(j-l) show, the average evolution age versus the node degree, for $W = 2, 4$ and 8, respectively. We observe an approximately monotonic relation for small degree. However, when the node degree is larger than 10, the relation deteriorates quickly and the relations approach a constant.

To further demonstrate our method, we have analyzed a scale-free cellular network generated by mechanism different from that of the preferential-attachment rule, namely the protein-protein interaction(PPI) networks. There is also another detailed work to describe the method of "Simulated Evolution of Protein-Protein Interaction Networks with Realistic Topology" published on "PLOS one 2012" [108]. In such a network, duplication and divergence are believed to be responsible for the topological structure [109].

We start from a small, connected graph as a seed and duplicate a randomly selected existing protein at each step. The new comer duplicates exactly the connection pattern of its generator in the network. Due to mutations, some of the

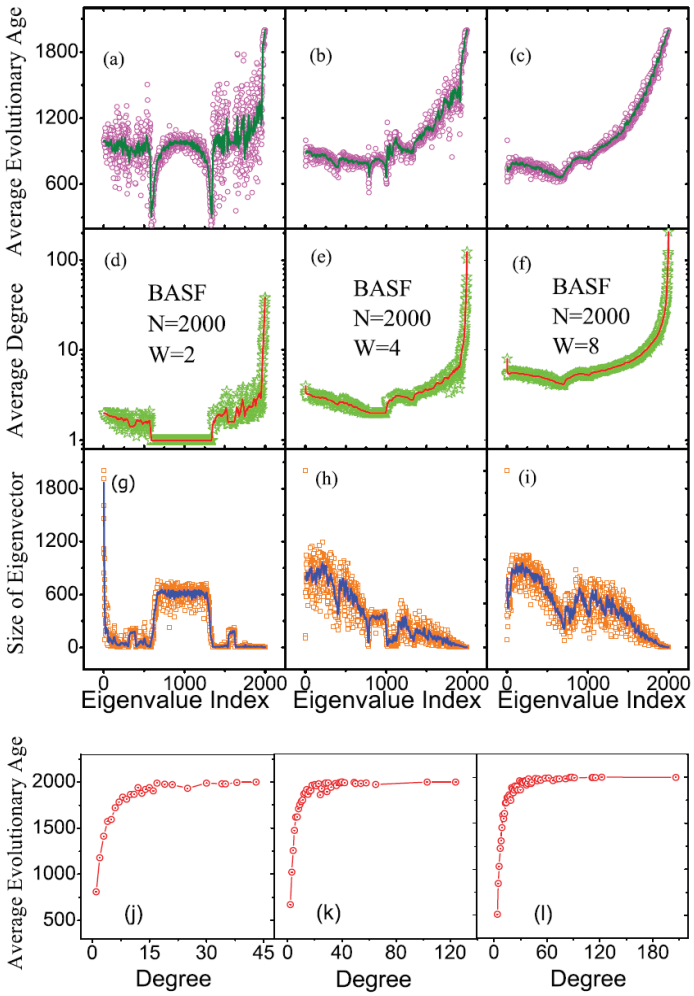

Figure 3.3: **The relation between eigenvalues and node ages for standard scale-free networks.** For three scale-free networks generated according to the standard preferential-attachment rule with edge density $w = 2, 4, 8$ (corresponding to the left, middle, and right column, respectively), (a-c) average ages, (d-f) average degree (on a logarithmic scale), and (g-i) size of eigenvector versus the eigenvalue index i. Eigenvectors associated with large eigenvalues generally have small sizes, but their ages are "older" in the network. (j-l) Average age versus degree. We see that, while small degree is related with the average age, information about node age deteriorates quickly as the degree is increased.

duplicated edges are broken with probability p, while new edges are generated with probability q between the new comer and other existing nodes.

To compare with the PPI network of the Baker's Yeast (to be described in the next Section), we generate networks with comparable parameters. In particular, a typical network has 2235 nodes and average degree of 10.52, and degree distribution follows power-law with exponent 2.3. In a wide range of eigenvalues there exists a strong correlation between the eigenvalue and average age, as shown in Fig.3.4(a). We observe that, the curve of average age versus degree exhibits large fluctuations, as shown in Fig.3.4(d). It is thus not possible to obtain information about node age from degree. However, behaviors of the eigenmodes can reveal the age information, as will be demonstrated in Sec. 3.4.

3.4 Evolution Ages of Nodes in a Protein-Protein Interaction Network

To lend more credence to our proposition that the evolutionary ages of nodes can be inferred from the eigenvalues, we now consider a class of networks in systems biology, protein-protein interaction (PPI) networks. These networks are the result of a number of evolutionary mechanisms such as duplications of genes and reattachments of links between the proteins. Specifically, we analyze the PPI network of the baker's yeast (*Saccharomyces cerevisiae*) [110, 111]. Von Mering et al. [112] analyzed a total of 80000 interactions among 5400 yeast proteins reported previously and assigned each interaction a confidence value. In order to reduce the effect of false positives, we focus on 11855 interactions with high and medium confidence values among 2617 yeast proteins. In a PPI network, each protein is a

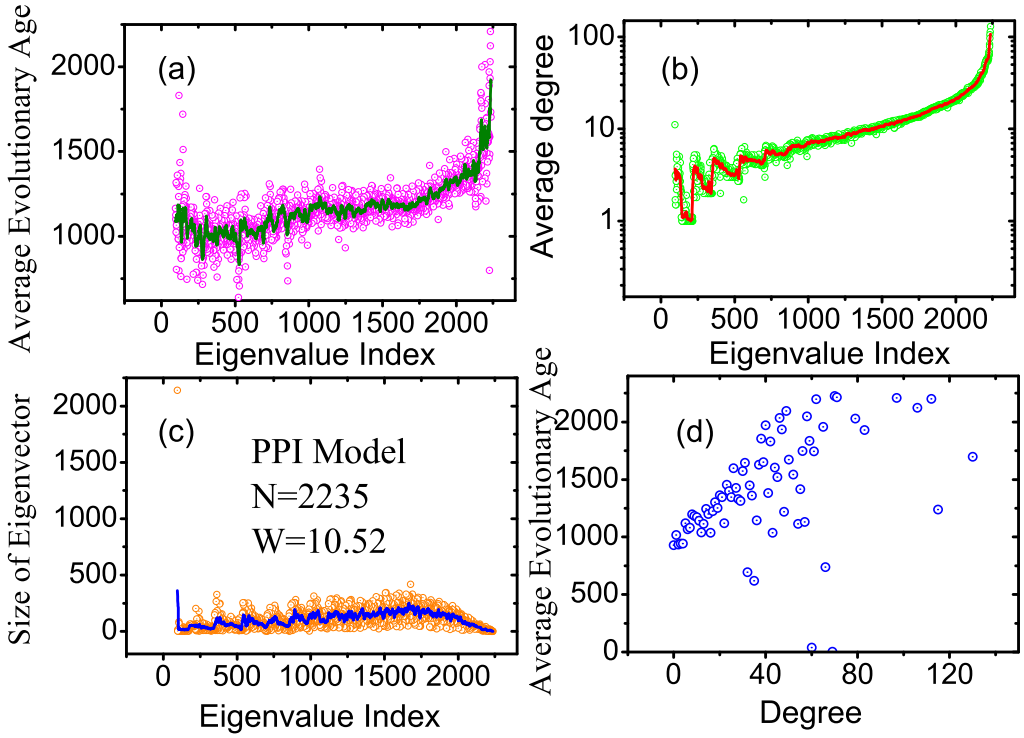

Figure 3.4: **The relation between eigenvalues and node ages using scale-free networks generated by duplication/divergence-based mechanism from PPI network of the Baker's Yeast.**For (a) average age versus eigenvalue index, (b) average degree versus eigenvalue index, and (c) size of eigenvector versus the eigenvalue index. Eigenvectors associated with large eigenvalues generally have small sizes, but their ages are "older" in the network. (d) Average age versus degree. Because of large fluctuation, the degree cannot give age-related information, but the eigenvalues can.

node and each pairwise interaction represents a link between two nodes. Since our goal is to assess, through the eigenvalues, the evolutionary ages of the nodes, we neglect the directions of the edges. The largest connected component of the PPI network contains 2235 nodes.

In systems biology, the evolutionary processes of the proteins are classified into four iso-temporal groups [113]: prokaryotes, eukarya, fungi, and yeast, to which numbers $4, 3, 2$ and 1 are assigned according to their evolutionary process from ancient to modern times, respectively. The evolutionary age of a protein is the largest number from the groups it presents. For example, the protein YHR037w occurs in the groups prokaryotes(4), eukarya(3), fungi(2), which means that it can be found from the ancient prokaryotes, so that its age is 4.

Figure 3.5(a) shows the average evolutionary age of nodes in eigenvector versus the eigenvalue index, which is similar to the behavior in Figs. 3.3(a-c). This suggests that for a realistic biological network, there is indeed a positive correlation between the eigenvalues of the Laplacian matrix and the evolutionary ages of groups of nodes. Since PPIs typically possess a scale-free structure [114], we expect the average degree of groups of nodes to exhibit similar behaviors as in Figs. 3.3(d-f). This is indeed the case, as shown in Fig. 3.5(b). The sizes of various eigenvectors are shown in Fig. 3.5(c). Again the behavior is similar to those in Figs. 3.3(g-i). From Fig. 3.5(d), relation of average age versus degree, we see that the degree contains no information about the node age.

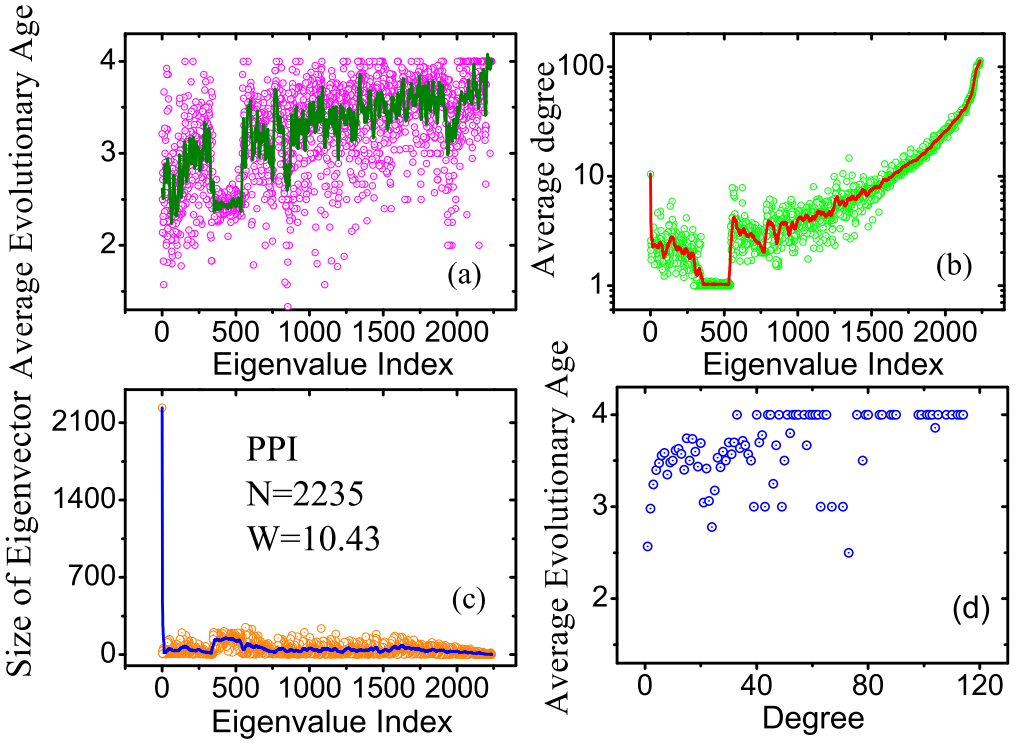

Figure 3.5: **The relation between eigevalues and node ages using the largest connected component of the real PPI network of the baker's yeast.** For (a) the evolutionary age, (b) average degree (on a logarithmic scale), and (c) size of eigenvector versus the eigenvalue index i. These results further indicate that the evolutionary ages of various nodes in the network can be inferred from the eigenvalue spectrum of the Laplacian matrix. (d) Average age versus degree. We see that degree does not reveal age-related information.

3.5 Time-Series Based Detection of Evolutionary Ages of Nodes

We now address the situation where the network topology is unknown but only time series measured or observed from various nodes are available. We shall apply a recently developed approach [100] based on compressive sensing [115–120] to uncover the complex-network topology and then could analyze the spectrum of the predicted Laplacian matrix to estimate the evolutionary ages of nodes. The unique feature of compressive sensing lies in its extremely low data requirement: very little observation is needed to obtain a target sparse signal. In the following section, the part for the compressive-sensing paradigm and the predicted eigenvalues in figure 3.7 are done by my co-authors Lai Ying-Cheng and Yang Rui from Arizona State University.

In general, the problem of compressive sensing can be described as to reconstruct a sparse vector $\mathbf{a} \in R^N$ from linear measurements $\mathbf{X}$ about $\mathbf{a}$ in the form: $\mathbf{X} = \mathbf{G} \cdot \mathbf{a}$, where $\mathbf{X} \in R^M$ and $\mathbf{G}$ is an $M \times N$ matrix. Accurate reconstruction can be achieved by solving the following convex optimization problem [115]

$$\min \|\mathbf{a}\|_1 \quad \text{subject to} \quad \mathbf{G} \cdot \mathbf{a} = \mathbf{X}, \tag{3.4}$$

where $\|\mathbf{a}\|_1 = \sum_{i=1}^{N} |\mathbf{a}_i|$ is the L_1 norm of vector $\mathbf{a}$ and $M \ll N$, *i.e.*,, the number of measurements can be much less than the number of components of the unknown signal. Various solutions of the convex optimization problem (3.4) have been worked out in the applied-mathematics literature [115–120].

To uncover network topology based on data, it is necessary to cast the problem in the form (3.4). The basic hypothesis is that a complex networked system can be

viewed as a large dynamical system that generates oscillatory time series at various nodes. Under this hypothesis, it is straightforward to formulate the problem under the compressive-sensing paradigm, details of which can be found in Ref. [100].

To give a concrete example, we consider a real-world network, the Santa Fe Institute (SFI) collaboration network [17]. Collaboration Networks, in which individuals are connected if they are (were) involved in a common activity, have been often studied because they introduced an implicit concept of acquaintance that is not easy to obtain in direct social experiments/interviews. The analysis of the structure of scientific collaboration networks [10] has triggered huge influences on the development of the modern network science. Scientific collaboration is a very important topic: two scientists are linked if they have co-authored at least one paper together. Information about co-authors hips can be extracted from different databases of research papers. Communities reflect groups of people with common research interests, i.e. topical or disciplinary groups.

Here, for the Santa Fe Institute (SFI) collaboration network [17], there are $N = 76$ nodes in the largest connected component of the network and the average degree is about 3. A schematic illustration of the network is shown in Fig. 3.6. A spectral analysis reveals that the eigenvectors associated with λ_{76}, λ_{75} and λ_{74} characterize the three hubs: 40, 7 and 67, all marked by red. The eigenvector associated with λ_{73} involves a group of nodes numbered between 17 and 25 (marked by green). For λ_{72}, the corresponding eigenvector covers nodes 26 to 29, and node 34 (marked by cyan). The three clusters: nodes 41 to 47 (blue), 1 to 6 (magenta), and 48 to 53 (violet), are represented by eigenvectors λ_{70}, λ_{69}, and λ_{68}, respectively. In fact, clusters of larger scales can be identified for smaller eigenvalues.

Now assume that the network topology is unknown but an oscillatory time

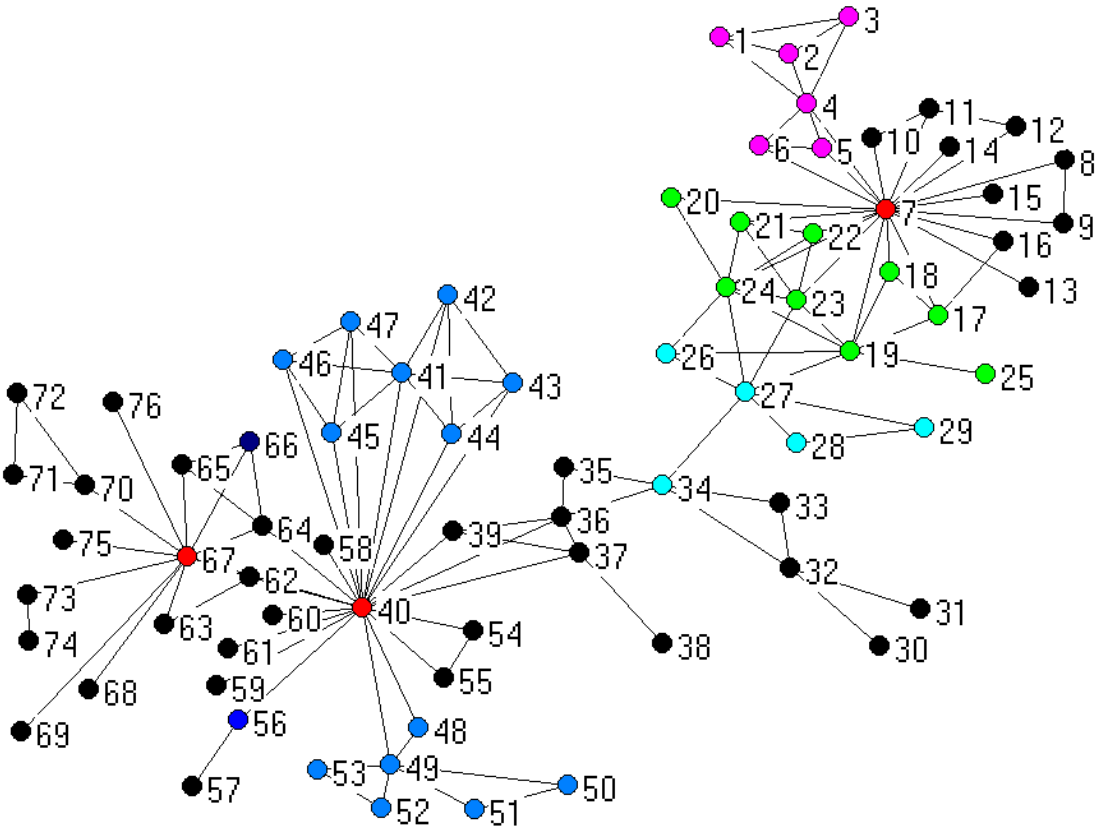

Figure 3.6: **Schematic illustration of the largest component of the SFI collaboration network and the clustered structure revealed by an eigenvalue/eigenvector analysis.**

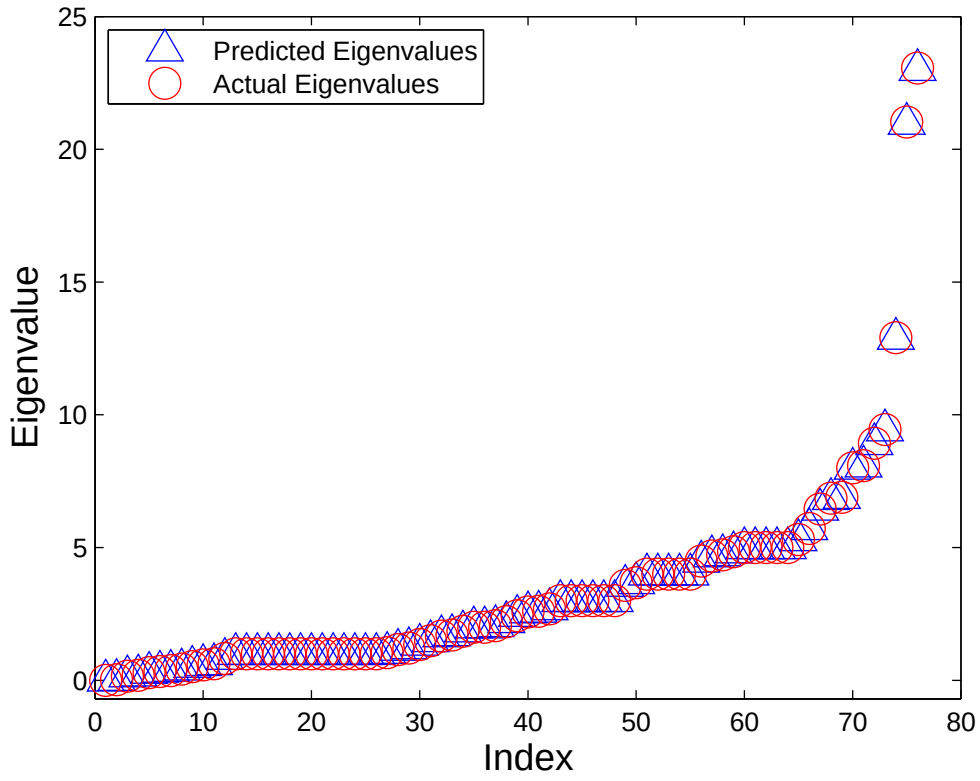

Figure 3.7: **Sorted eigenvalues of the predicted and actual Laplacian matrix of the SFI collaboration network.** The number of data points used in uncovering the network structure is about 40% of the number of total unknown coefficients in the power-series expansion.

series from each node is available. To simulate the situation, we assume that the dynamics of each node is described by the chaotic Rössler oscillator [121]. Applying the compressive-sensing based method to uncover the network topology, we can then perform a spectral analysis to estimate the ages of various nodes in the network. Figure 3.7 shows the eigenvalues of the predicted and the actual Laplacian matrix. We observe an excellent agreement.

3.6 Summary

In summary, we have developed a spectral-analysis based method to estimate the evolutionary ages of nodes in complex networks. The basic observation is that eigenvectors associated with different eigenvalues of the Laplacian matrix can typically represent highly localized groups of nodes in the network. In particular, it was found that there are complex networks in the real-world for which there is a positive correlation between the eigenvalue magnitude and node age.

Knowledge of ages of various groups of nodes can provide significant insights into the evolutionary process underpinning the network. It should be noted, however, that at the present the applicability of our method is limited to the networks for which information about the node age has been encoded gradually in the eigenproperties through evolution.

Chapter 4

Conclusions and Future Perspectives

This dissertation has presented a theoretical study of how to use dynamic-based structure measures to characterize complex network, namely, by means of their eigenvalue spectra and especially their localized eigenvectors. We have also demonstrated that both the undirected and directed complex networks have incredible localization properties. Moreover, the study also leads to some significant insights into the evolutionary process underpinning the networks.

First, we have analyzed spectra and localization properties of undirected networks. The probability distribution function of the representative eigenvector is proposed to describe the localization properties of complex networks here. The localization quantities (Q, S_{str}), the PDFs of the NNLS, and the wavelet transform are used to capture the characteristics of the representative state. The nontrivial structures of the networks can induce the localizations of the representative states.

At the same time, because of the global symmetries of the networks, the representative states have nontrivial structures rather than the step-like distribution.

The localization quantities (Q, S_{str}) and Brody distribution parameter β can describe the nontrivial localization properties in a global way. The (Q, S_{str}) values tell us that the BASF networks with $w = 2$ are significantly localized compared with the WSSW networks with $d = 2$. It is consistent with the conclusions drawn from the results of β. The whole cellular networks have localization properties much closer to the WSSW networks.

The wavelet transform can tell us the details on the nontrivial structures of the representative eigenvectors. The ascend-ranked series ρ for the WSSW and BASF modeling networks and the whole cellular networks behave mono-fractal, multi-fractal or branched multi-fractal. The PDF of ρ tends to form sharp peaks at different scales in a self-similar way.

These kinds of properties can shed light on the global symmetries due to the general rules in the construction of the networks. Hence, it can be employed as a global measure of the network structures. Moreover, the structure-induced localizations on networks maybe helpful to understand the electronic conduction and heat transport properties [122] of nanonet materials.

In additional to our work, Shepelyansky' group from France has also applied the spectra analysis method to the research of the localization properties of eigenvectors for the Google matrix and the BA scale free model network. They found the emergence of a delocalization phase for the PageRank vector in google matrix when network parameters are changed. Their results show that actual WWW networks are located in the localized phase. The transition to the delocalized phase can drastically affect the efficiency of the Google search [92].

A closely relevant topic is the diffusion on complex networks. Kim et al. [123] reported for the first time their works on quantum and classical diffusion on WSSW networks. The Hamiltonian is same as that in the present thesis, namely $\varepsilon_n = 0$ and $t_{mn} = 1$ in Eq. 2.1. An electron is localized at a randomly selected node at beginning, then the diffusion process is obtained by solving the time-dependent Schrodinger equation. It is found that the "long-range" edges can boost the diffusion speeds significantly, especially at the transition point from $p_r = 0$ to $p_r \neq 0$. This is qualitatively in consistent with our findings of the significant changes of the participation ratio and the structural entropy, (Q, S_{str}), when p_r increase from $p_r = 0$ to $p_r = 0.02$.

As for the classical diffusion on networks, a very recent work reports the first-passage times (FPT) of random walkers in complex scale-invariant media [13, 124]. Many real-world networks have self-similar structures, and diffusion on networks can be regarded, to a certain degree, as the diffusion on fractal media, which has attracted intensive attentions for its importance in theories and potential use in diverse research fields [125].

However, we should point out that, it is not trivial to compare our results quantitatively with these evolution processes. Actually our results are obtained from the eigenstates in energy representation, while for the quantum diffusion the initial state localizing at a randomly selected node is a wave packet and the final state should be a superposition of the eigenstates in energy representation. How to relate the localization with the classical diffusion is definitely interesting but not a trivial task. Obviously, detailed works on diffusion on complex networks are required to understand the relation between localization and diffusion on networks.

In addition to the study of localization properties of complex networks, we

have also investigated the spectral properties of random networks on directed networks. The network edges take value $+1$ and -1 depending upon whether it is starting from an exhibitory node or from an inhibitory node. If all nodes are excitatory then the corresponding network is symmetric ($\tau = 1$). Directionality is introduced by making some nodes inhibitory, and consequently corresponding edges take value -1. Equal expected value of inhibitory and excitatory nodes gives rise to a completely un-correlated network ($\tau = 0$).

Spectra of random networks, where probability for a node being inhibitory or excitatory is equal, show circular distribution with radius being $\sqrt{pN(1-p)}$. Based on IPR values the spectra can be divided into two parts: part A consisting of eigenstates at real axis and at four corners with large absolute eigenvalues which are localized, and part B consisting of the bulk middle part of the spectra which is less localized. As connection probability increases, part B starts dominating the spectra, except few localized eigenvalues which remain very well separated from the bulk part even for very large connection probabilities.

Moreover, in order to understand the mechanism for localization, we track the spectra as network is rewired from a completely symmetric structure ($\tau = 1$) to a completely asymmetric one ($\tau = 0$). For symmetric networks spectra lie on real axis with exactly one eigenvalue separated from the bulk. As connections are made directed by making some nodes inhibitory, some of eigenstates start occurring in complex conjugate pairs. The eigenvalue distribution along with IPR value show rich pattern. Overall, the spectra gradually become more delocalized as number of directed connections is increased.

We have investigated spectra and localization properties of directed networks with binary entries. The networks with inhibitory and excitatory nodes have much

richer spectra than the networks with only excitatory nodes. Bulk of the spectra for completely asymmetric networks follow Girko's law, but as probability of directed connections is reduced the spectra show very different patterns depending upon the network structure and ratio of inhibitory and excitatory nodes. Though directed networks span varieties of complex systems, the research for directed networks leading to complex eigenvalues is limited. The results presented in the thesis provide a useful platform to understand the structural pattern in directed networks, and can be used further to investigate dynamical behavior of nodes relevant to variety of problems ranging from physics to sociology.

Finally, we move to develop a procedure using spectral-analysis based method to estimate the evolutionary ages of nodes in complex networks. The basic observation is that eigenvectors associated with different eigenvalues of the Laplacian matrix can typically represent highly localized groups of nodes in the network.

A qualitative argument can then be made for the existence of positive correlation between the node ages and the magnitudes of the eigenvalues. This means that, when the network topology is known, a simple eigenvalue analysis can lead to reliable information about the age distribution of nodes in the network.

For situations where the network topology is unknown but time series from nodes are available, it is necessary to uncover the topology in order to estimate the node ages, and we have demonstrated that this can be done efficiently using compressive sensing. Examples from model and real-world networks, including a PPI network, are used to validate our approach.

We hope our method can find applications in fields such as systems biology, the propagation of a rumor, a fashion, a joke, or a flu, where estimating node ages can be of significant important.

Indeed some progress have been achieved. Most recently, an Italian research group [126] applied our methods to predict the sources of an outbreak by testing on a variety of graphs collected from outbreaks including influenza, H5N1, Tbc, in urban and rural areas. Results show that the spectral analysis method is able to identify the source nodes if the graph approximates a tree sufficiently.

The network-reconstruction technique used in the thesis is based on compressive sensing, which works for situations where the types of mathematical forms of the nodal dynamical systems and coupling functions are known (although details of these functions are not required) and can be represented by series expansion. So far the method has not been applied to gene-regulatory networks due to difficulty to find suitable series expansions. The recent method by Hempel et al. [127] is based on extracting statistical information and has been demonstrated to work well for gene-regulatory networks.

While many real-world systems such as gene regulatory and supply chain networks are directed, our present work is focused on undirected networks. The main consideration is that many networks generated by some kind of evolutionary processes or constructed through experiments tend to undirected. For example, the Baker Yeast obtained through the approach of prey and predator contains no information about the directionality of the nodal interactions.

Our method is based on the observation that local structures, e.g., densely connected clusters, can induce large components in the eigenvectors. Hubs or clusters of hubs can then be detected by the eigenvectors corresponding to the largest eigenvalues, while clusters of larger sizes can be uncovered by eigenvectors of smaller eigenvalues. Different eigenmodes can be used to detect clusters of varying scales, providing a correlation with the evolutionary ages in situations

where hubs or clusters of hubs are formed by history. The principle on which our method is based thus does not take into account directionality in the node-to-node interactions. To develop a method to uncover the evolutionary ages for directed complex networks remains to be an interesting but open question at the present.

Bibliography

[1] D. J. Watts, Small Worlds (Princeton University Press, Princeton, 1999); R. Albert and A. -L. Barabasi, Rev. Mod. Phys.**74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, Evolution of Networks (Oxford University Press, New York, 2003).

[2] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[3] A. -L. Barabasi and R. Albert, Science **286**, 509 (1999).

[4] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[5] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabasi, Science **297**, 1551 (2002).

[6] C. Song, S. Havlin, and H. A. Makse, Nature **433**, 6392 (2005); C. Song, S. Havlin, and H. A. Makse, Nature Physics **2**, 275 (2006).

[7] L. K. Gallos, C. Song, S. Havlin, and H. A. Makse, Proc. Natl. Acad. Sci. USA **104**, 7746 (2007).

[8] H. Yang, C. Yin, G. Zhu, and B. Li, Phys. Rev. E **77**, 045101**(R)** (2008).

[9] B. J. Kim, Phys. Rev. Lett. **93**, 168701 (2004).

[10] M. E. J. Newman, SIAM Review **45**, 167 (2003).

[11] C. Zhou and J. Kurths, Chaos **16**, 015104 (2006); K. Park, Y. Lai, and S. Gupte, Chaos **16**, 015105 (2006).

[12] V. Sood and P. Grassberger, Phys. Rev. Lett. **99**, 098701 (2007).

[13] L. D. F. Costa, and G. Travieso, Phys. Rev. E **75**,016102 (2007).

[14] A. S. Ribeiro, S. A. Kauffman, J. Lloyd-Price, B. Samuelsson, and J. E. S. Socolar, Phys. Rev. E **77**,011901 (2008); P. Krawitz and I. Shmulevich, Phys. Rev. E **76**,036115 (2007).

[15] M. L. Mehta, Random Matrices, (3rd ed. Elsevier Academic, Amsterdam, 2004).

[16] N. D. Martinez, Artifacts or attributes? Effects of resolution on the Little Rock Lake food web, Ecological Monographs **61**, 367-392 (1991).

[17] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U. S. A. **99**, 7821 (2002).

[18] J. J. Potterat, Phillips-Plummer, L., Muth, S. Q., Rothenberg, R. B., Woodhouse, D. E., MaldonadoLong, T. S., Zimmerman, H. P., and Muth, J. B., Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs, Sexually Transmitted Infections **78**, i159-i163 (2002).

[19] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51** (4), 1079-1187 (2002).

[20] S. Havlin and R. Cohen, Cambridge University Press (2010).

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D. Hwang, Physics Reports **424** (4-5), 175-308 (2006).

[22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, Science **298** (5594), 824-827 (2002).

[23] S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon, Nature Genet. **31** (1), 64-68 (2002).

[24] S. Fortunato, Physics Reports-Review Section of Physics Letters **486** (3-5), 75-174 (2010).

[25] H. A. Simon, American Economic Review **52** (2), 1-15 (1962).

[26] J. C. Chen and B. Yuan, Bioinformatics **22** (18), 2283-2290 (2006).

[27] J. Scott, British Journal of Sociology **51** (4), 754-755 (2000).

[28] V. D. Blondel, J. L. Guillaume, R. Lambiotte and E. Lefebvre, Journal of Statistical Mechanics-Theory and Experiment (2008).

[29] J. Davidsen, H. Ebel, S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).

[30] R. Monasson, Eur. Phys. J. B **12**, 555 (1999).

[31] M.E.J. Newman, D.J. Watts, Phys. Lett. A **263**, 341 (1999).

[32] D.J.de S. Price, J. Amer. Soc. Inform. Sci. **27** 292(1976).

[33] D.J.de S. Price, Science **149** 510(1965).

[34] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Phys. Rev. Lett. **85** (21), 4626-4628 (2000).

[35] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Phys. Rev. Lett. **86** (16), 3682-3685 (2001).

[36] D. S. Callaway, M. E. J. Newman, S. H. Strogatz and D. J. Watts, Phys. Rev. Lett. **85** (25), 5468-5471 (2000).

[37] J. J. Ramasco, and M. Mungan, Phys. Rev. E **77**(3), 036122 (2008).

[38] A.E. Motter, C. Zhou, J. Kurths, Phys. Rev. E, **71**, 016116 (2005).

[39] T. Guhr, A. Muller-Groeling, and H. A. Weidenmuller, Phys. Rep. **299**, 189 (1998).

[40] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).

[41] P. Seba, Phys. Rev. Lett. **91**, 198104 (2003).

[42] M. S. Santhanam and P. K. Patra, Phys. Rev. E **64**, 016102 (2001).

[43] D. M. Cvetkovic, M. Doob and H. Sachs, Academic Press, 3rd Revised edition, (1997).

[44] F.R.K. Chung, Spectral Graph Theory (American Mathematical Society, Providence, Rhode Island, (1997).

[45] L.M. Pecora and T.L. Carroll, Phys. Rev. Lett.**80**, 2109 (1998).

[46] K.S. Fink, G. Johnson, T.L. Carroll, L.M. Pecora, Phys. Rev. E **61**, 5080 (2000).

[47] C. Zhou, A. E. Motter and J. Kurths, Phys. Rev. Lett. **96** 034101 (2006).

[48] S. Boccaletti, J. Kurths, D.L. Valladares, G. Osipov, C.S. Zhou, Phys. Rep. **366**, 1 (2002).

[49] R. Grone, R. Merris, and V. S. Sunder, The Laplacean spectrum of a graph, SIAM J. Matrix Analysis and Appl. **11**, 218 - 238 (1990).

[50] A. Banerjee and J. Jost, Theory in Biosciences **126**, 15-21 (2007).

[51] I. J. Farkas et al., Phys. Rev. E **64**, 026704, (2001).

[52] S. N. Dorogovtsev et al., Phys. Rev. E **68**, 046109 (2003).

[53] S. Jalan and J. N. Bandyopadhyay, Phys. Rev. E **76**, 046107 (2007).

[54] S. Jalan and J. N. Bandyopadhyay, Physica A **387**, 667-674 (2008).

[55] S. Jalan, N. Solymosi, G. Vattay and B. Li, Phys. Rev. E **81**, 046118 (2010).

[56] G. M. Zhu, H. J. Yang, C. Y. Yin, B. Li, Phys. Rev. E **77**, 066113 (2008).

[57] S. Jalan, G. M. Zhu, and B. Li, Phys. Rev. E **84**, 046107 (2011).

[58] H. Yang, F. Zhao, L. Qi, and B. Hu, Phys. Rev. E **69**,066104 (2004); F. Zhao, H. Yang, and B. Wang, Phys. Rev. E **72**,046119 (2005); H. Yang, F. Zhao, and B. Wang, Physica A **364**,544 (2006); H. Yang, F. Zhao, and B. Wang, Chaos **16**,043112 (2006).

[59] S. Jalan, and J. N. Bandyopadhyay, Euro. Phys. Letts **87**, 48010 (2009).

[60] H. Yang, F. Zhao, and B. Wang, Physica A **364**,544 (2006);

[61] C. Ellegaard, T. Guhr, K. Lindemann, J. Nygard and M. Oxborrow, Phys. Rev. Lett. **77**, 4918 (1996).

[62] T. Gross, C. J. Dommar DLima, and Bernd Blasius, Phys. Rev. Lett. **96**, 208701 (2006)

[63] B. Pierre, F. Eric, et. al, Proceedings of NATO Advanced Study Institute on Mining Massive Data Sets for Security, IOS Press, (2008).

[64] G. Pall, A. Barabasi, and T. Vicsek, Nature **446**, 664-667 (2007).

[65] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. E **64**,051903 (2001); S. N. Dorogovtsev, A. V. Goltsev, J. F. Mendes, and A. N. Samukhin, Phys. Rev. E **68**, 046109 (2003); C. Zhu, S. Xiong, Y. Tian, N. Li and K. Jiang, Phys. Rev. Lett. **92**, 218702 (2004); C. Kamp, and K. Christensen, Phys. Rev. E **71**, 041911 (2005); M. Sade, T. Kalisky, S. Havlin, and R. BerkovitsPhys. Rev. E **72**, 066123 (2005); P. N. McGraw, and M. Menzinger, Phys. Rev. E **75**, 027104 (2007); J. N. Bandyopadhyay and S. Jalan, Phys. Rev. E **76**, 026109 (2007).

[66] S. Xiong and S.N. Evangelou, Phys. Rev. B **52**, 13079(R) (1995); C. Zhu, and S. Xiong, Phys. Rev. B **62**, 14780 (2000); L. Gong, and P. Tong, Phys. Rev. E **74**, 056103 (2006).

[67] J. Yi, and B. J. Kim, Phys. Rev. B, **76**, 245207 (2007).

[68] M. Dresselhaus, G. Dresselhaus,P. Eklund, and R. Saito, Phys. World **11**, 33 (1998); P. G. Collins, and P. Avouris, Sci. Am. **283**, 62 (2000); R. H. Baughman, A. A. Zakhidov, and W. A. de Heer, Science **297**, 787 (2002); G. Gruner, J. Mat. Chem. **16**, 3533 (2006);G. Gruner, Anal. Bio. Chem. **384**, 322 (2006); G. Gruner, Sci. Am. **17**, 48 (2007).

[69] E. Lopez, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **94**, 248701 (2005); Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **96**, 148702 (2006); G. Li, L. A. Braunstein, S. V. Buldyrev, S. Havlin, and H. E. Stanley, Phys. Rev. E **75**, R045103 (2007).

[70] P. W. Anderson, Phys. Rev. **109**, 1492 (1958); P. A. Lee, and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).

[71] Y.-J. Kim, M. H. Lee, and M. Y. Choi, Phys. Rev. B **40**, 2581 (1989).

[72] M. Titov, and H. Schomerus, Phys. Rev. Lett. **91**, 176601 (2003).

[73] C. Tang, and M. Kohmoto, Phye. Rev. B **34**, 2041 (1986).

[74] M. Kohmoto, B. Sutherland, and C. Tang, Phys. Rev. B **35**, 1020 (1987).

[75] M. Mezard, G. Parisi, and A. Zee, Nucl. Phys. B **559**, 689 (1999); S. Ciliberti, T. S. Grigera, V. Martn-Mayor, G. Parisi, and P. Verrocchio, Phys. Rev. B **71**, 153104 (2005).

[76] M. Sade, and R. Berkovits, Phys. Rev. B **68**, 193102 (2003).

[77] E. L. Albuquerque and M. G. Cottam, Phys. Rep. **376**, 225 (2003).

[78] M. Janssen, Phys. Rep. **295**, 1 (1998).

[79] A.D. Mirlin, Phys. Rep. **326**, 259 (2000).

[80] J. Pipek and I. Varga, Phys. Rev. A **46**, 3148 (1992).

[81] J. Pipek, Phys. Rev. E **68**, 026202 (2003); and the references there-in.

[82] I. Varga, Phys. Rev. B **66**, 094201 (2002).

[83] T. Dittrich, Phys. Rep. **271**, 267 (1996).

[84] R. Berkovits and Y. Avishai, Phys. Rev. B **53**, R16125 (1996).

[85] B. L. Altshuler, I. Kh. Zharekeshev, S. A. Kotochigova, and B. I. Shklovskii, Zh. Eksp. Teor. Fiz **94**, 343 (1988) [Sov. Phys. JETP **67**, 625 (1988).

[86] B. I. Shklovskii, B. Shapiro, B. R. Scars, P. Lambrianides, and H. B. Shore, Phys. Rev. B **47**, 11487 (1993).

[87] E. Hofstetter and M. Schreiber, Phys. Rev. B **48**, 16979 (1993).

[88] M.L. Goldstein, S.A. Morris, and G.G. Yen, Eur.Phys. J. B **41**, 255 (2004); M. E. J. Newman, Contemp. Phys. **46**, 323 (2005);

[89] P. Ch. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature **399**, 461 (1999).

[90] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai,and A.-L. Barabasi,Nature **407**, 651 (2000).

[91] R. Overbeek, et al., Nucleic Acid Res.**28**, 123 (2000).

[92] O. Giraud, B. Georgeot and D. L. Shepelyansky, Phys. Rev. E **80**, 026107 (2009)

[93] B. Georgeot, O. Giraud, and D. L. Shepelyansky Phys. Rev. E **81**, 056109 (2010).

[94] K. Rajan and L. F. Abbott , Phys. Rev. Lett. **97**, 188104 (2006).

[95] O. Hul, P. Šeba and L Sirko, Phys. Rev. E **79** 066204 (2009).

[96] P. Erdös and A. Rényi, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).

[97] D. W. Lang, Physical Review, **135**, B1082 (1964).

[98] G.-M. Zhu, H.J. Yang, R. Yang, J. Ren, B. Li, and Y.-C. Lai, Eur. Phys. J. B. **85**, 106 (2012).

[99] M. J. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).

[100] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and M. A. F. Harrison, Europhys. Lett. **94**, 48006 (2011); W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Phys. Rev. Lett. **106**, 154101 (2011).

[101] H. J. Yang, F. C. Zhao, and B. H. Wang, Chaos **16** (2006).

[102] J. Ren, and B. Li, Phys. Rev. E **79**, 051922 (2009).

[103] J. Ren, W. X. Wang, B. Li, and Y. C. Lai, Phys. Rev. Lett. **104**, 058701 (2010).

[104] J. G. Restrepo, E. Ott, and B. R. Hunt, Phys. Rev. Lett. **93**, 114101 (2004).

[105] K. Park, L. Huang, and Y.-C. Lai, Phys. Rev. E **75**, 026211 (2007).

[106] I. Bahar, A. R. Atilgan, and B. Erman, Folding & Design **2**, 173 (1997).

[107] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).

[108] Peterson, G. Jack, Presse, Steve, Peterson, Kristin S., et al. PLOS ONE Volume: 7 Issue: 6 , e39052 (2012)

[109] A. V. Vazquez, A. Flammini, A. Maritan, A. Vespignani, Complexus **1**, 38 (2003).

[110] A. Wagner, Mol. Biol. Evol. **18**, 1283 (2001).

[111] A. Wagner, Proc. R. Soc. Lond. Ser. B-Biol. Sci. **270**, 457 (2003).

[112] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork, Nature **417**, 399 (2002).

[113] C. R. Woese, Microbiol. Rev. **51**, 221 (1987).

[114] E. Ravasz, A. L. Somera, D. A. Mongru, Z. Oltvai, and A.-L. Barabási, Science **297**, 1551 (2002).

[115] E. Candès, J. Romberg, and T. Tao, IEEE Trans. Inf. Theory **52**, 489 (2006); Commun. Pure Appl. Math. **59**, 1207 (2006).

[116] E. Candès, in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.

[117] D. Donoho, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[118] R.G. Baraniuk, IEEE Signal Processing Mag. **24**, 118 (2007).

[119] E. Candès and M. Wakin, IEEE Signal Processing Mag. **25**, 21 (2008).

[120] E. Candès, and J. Romberg, http://www. acm. caltech.edu/l1magic, 2005.

[121] O. E. Rössler, Phys. Lett. A **57**, 397 (1976).

[122] Z. Liu, and B. Li, Phys. Rev. E**76**, 051118 (2007).

[123] B. J. Kim, H. Hong, and M. Y. Choi, Phys. Rev. B **68**, 014304 (2003).

[124] S. Condamin, O. Benichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature **450**, 77 (2007).

[125] F. D. A. Aarao Reis, J. Phys. A **29** 7803 (1996); C. Schulzky, A. Franz, and K. H. Hoffmann, SIGSAM Bull. **34**, 1 (2000); A. Franz, C. Schulzky, S. Tarafdar, and K. H. Hoffmann, J. Phys. A **34**, 8751 (2001); D. H. N. Anh, K. H. Hoffmann, S. Seeger, and S. Tarafdar, Europhys. Lett. **70**, 109 (2005);D. H. N. Anh, P. Blaudeck, K. H. Hoffmann, J Prehl and S. Tarafdar, J. Phys. A **40**, 11453 (2007).

[126] V Fioriti, M Chinnici, arXiv preprint arXiv:1211.2333, 2012.

[127] S. Hempel, A. Koseska, J. Kurths, and Z. Nikoloski, Phys. Rev. Lett. **107**, 054101 (2011).